Journal of
Experimental Psychology

VOL. 25, NO. 4 OCTOBER, 1939

SENSORY PRE-CONDITIONING

BY W. J. BROGDEN

Pavlovian Laboratory, Phipps Psychiatric Clinic,
Johns Hopkins University School of Medicine

INTRODUCTION

The present investigation was designed to answer the following question: if an organism be given successive experiences of two temporally simultaneous stimuli exciting two sense modalities without evoking any observable response, and if after this contiguous sensory experience, one stimulus be made a conditioned signal for the activity of a given behavior-system by appropriate training, will the other elicit a similar conditioned response without the usual training?

Several studies have been reported which bear closely upon this problem. Prokofiev and Zeliony (**4**) presented human subjects with the sound of a metronome for 10 seconds and immediately afterwards gave a rythmical pressure on the skin of the fore-arm for an additional 10 seconds. This procedure was followed at intervals of from 3 to 15 minutes, from 5 to 10 times per day for several days. In addition, a whistle was blown for 5 to 10 seconds from time to time. The tactile stimulus was then presented without the metronome and followed by an electric shock to the hand. Within a few trials, reflex retrogression of the arm became conditioned to the tactile stimulus. When this conditioned response had been firmly established, the metronome was now presented alone. Two of the three subjects responded to the metronome with

retrogression of the arm. None of the subjects made any response to the whistle. In the same paper, these authors report the following experiment with a dog. Salivation, evoked by the introduction of hydrochloric acid into the mouth, was conditioned to the sound of a metronome. With this conditioned response firmly established, the introduction of acid into the mouth (without the sounding of the metronome) was then followed immediately by an electric shock to the foot, evoking flexion. When this combination had been presented a sufficient number of times to establish firmly conditioned flexion to the introduction of acid into the mouth, the metronome was given. To this stimulus, the animal salivated, but made no foot withdrawal.

Unfortunately, this investigation does not give a complete answer to the question proposed. There was no control over generalization of response. So, it is possible that human subjects conditioned to reflex retrogression of the arm to a tactile stimulus, without any prior experience of the tactile stimulus being associated with the sounding of the metronome, would respond to the sound of the metronome with arm-withdrawal. The long duration of the stimuli might tend to predispose the subject to respond to the stimulus not associated with shock. Furthermore, does response to the stimulus not associated with shock occur because the sound of the metronome functions as the equivalent of the tactile stimulus, or because the subject expects the tactile stimulus to follow the metronome?

Shipley (**5**) conditioned the eye wink, evoked by a sudden tap on the cheek, to the faint flash of an electric light. The tap on the cheek was now paired with shock to the finger, until it became an effective stimulus for finger-withdrawal. The subsequent presentation of the flash produced sudden withdrawals of the finger in 9 out of 15 subjects. When the tap on the cheek was given alone, then paired with shock, the presentation of the light did not evoke finger-withdrawal in any of 10 subjects. In a second control group the light and tap on the cheek were first associated; then shock to the finger was given alone. One of the 11 subjects in this group re-

sponded to the light by finger-flexion. Later, Shipley (**6**) put two groups of 10 subjects each through an experimental program consisting of two discrete conditioning steps. In the first step the eye wink was conditioned to a buzz and a light with one group; the buzz alone was used with the other group. In the second step, withdrawal of the finger was conditioned to the flash of a light. When the buzz was presented alone, it evoked finger-withdrawal in the subjects of both groups, the percentage of subjects responding being 60 and 70 respectively. Two control groups reveal that finger-withdrawal did not occur at the test point unless the test stimulus was one that had been previously used in one of the two steps in the experiment; and that finger-withdrawal occurred in but a small percentage of subjects when the shock employed in the second step had not been paired with an extra stimulus.

Shipley's experiments do not answer our question, for as in the case of the experiment of Prokofiev and Zeliony, the test-stimulus preceded the stimulus through which the association was made. Light and buzz preceded the tap to the cheek, just as the sound of the metronome preceded the tactile stimulus. Shipley's results are further complicated for an answer to the present problem. Light, buzz, and tap to the cheek all produced a definite motor response, an eye wink. We seek evidence of a relationship between sensory stimuli, evoking no phasic activity.

Our problem is stated specifically as follows: if dogs be given successive stimulations of sound and light in combination contiguously, and if either sound or light be made a conditioned signal for forelimb flexion from an electrically-charged grid, will the stimulus not involved in conditioning evoke flexion without being combined with shock?

Experimental Procedure

To answer this question, eight mongrel dogs, unselected in any manner, were placed in the stock of a sound-proofed, light-shielded experimental chamber and presented with the sound of a bell and the flash of a light in combination for 2 seconds. The sound was produced by an electric doorbell and the flash of light by a 150 watt unfrosted electric lamp. Twenty such stimulus combinations were given daily for

10 successive days. The eight animals were then arbitrarily divided into two groups of four animals each. Left forelimb flexion, with shock the unconditioned stimulus, was conditioned to bell in one group (group BE). The bell was 2 seconds in duration and was followed immediately by the shock which lasted for 1/10 of a second. The second group (group LE) was conditioned to light in a similar manner. Conditioning proceeded normally, with 20 trials per day as the test-unit; the conditioned stimulus (either light or bell, as the case might be) evoking flexion with subsequent shock-avoidance more and more frequently. The day after conditioned flexion reached 100 percent (20 shock-avoiding flexion-responses in one test-period), group BE (conditioned to bell) was given 20 trials per day of the light alone, until a score of zero-flexion was reached. Group LE (conditioned to light) was given bell alone until flexion to bell was extinguished to zero. In order to check on generalization, two control groups of four animals each were trained. With one group (group BC), forelimb flexion was conditioned to bell. When 100% flexion to bell was attained, the light was presented alone. With the other group (group LC), forelimb flexion was conditioned to light and when conditioning reached the 100 percent level, bell was given alone. Needless to say, neither of these two groups were given the experience of bell and light in combination, prior to the conditioning procedure.

Quantitative Results

The conditioning of one of the two stimuli did make the other an effective stimulus for flexion. The four animals in group BE, conditioned to bell, responded to light a total of 27 times in the 11 test-periods required to reach zero-response (Table I). Animals No. 2 and No. 4 made no response to light. The control animals for this group, group BC, made no response to light in a total of 4 test-periods (Table II). Group LE, conditioned to light, responded to bell a total of 56 times in the 16 test-periods needed to produce zero-response (Table I). All of the animals in this group made some response to bell. The control animals for this group, group LC, responded to bell 4 times in 5 test-periods (Table II). Animal No. 15 alone showed generalization by responding to bell. The experimental animals, grouped together, responded to the stimulus never associated with shock a total of 78 times, whereas the control animals, similarly grouped, responded to the stimulus never associated with shock nor with the conditioned stimulus only 4 times. The mean frequency of response for the experimental group is 9.75 (S.E. = 4.17) and the mean frequency of response for the control group is 0.50 (S.E. = 0.50). The mean difference of 9.25 flexion-responses (S.E. = 4.20 flexion-responses), when subjected to Fisher's test for the significance of differences between small samples,

TABLE I

RESULTS OBTAINED FROM THE EXPERIMENTAL ANIMALS

The percent scores represent the number of times conditioned flexion occurred to the conditioned stimulus. Twenty stimulus-presentations were given during each test-period.

	Group BE				Group LE			
Animal.....	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8
Bell and light given in combination for 2 seconds, 20 times per day for 10 days								
	Bell + shock to left forepaw				Light + shock to left forepaw			
1	10%	0%	0%	15%	0%	0%	10%	0%
2	0%	0%	0%	20%	0%	0%	0%	15%
3	0%	50%	0%	20%	20%	35%	15%	10%
4	35%	85%	50%	30%	20%	25%	0%	0%
5	90%	85%	75%	40%	75%	15%	0%	0%
6	95%	95%	90%	75%	50%	20%	35%	0%
7	95%	95%	95%	85%	55%	30%	70%	15%
8	90%	95%	75%	75%	65%	10%	55%	10%
9	90%	95%	95%	70%	95%	55%	75%	40%
10	100%	95%	95%	80%	100%	60%	90%	60%
11		100%	90%	95%		30%	90%	75%
12			95%	90%		65%	85%	90%
13			100%	95%		70%	100%	85%
14				100%		95%		90%
15						90%		90%
16						85%		100%
17						90%		
18						100%		
Critical Tests								
	Light alone				Bell alone			
1	55%	0%	15%	0%	40%	25%	10%	20%
2	30%		10%		5%	0%	0%	20%
3	10%		10%		5%			45%
4	5%		0%		0%			20%
5	0%							30%
6								25%
7								10%
8								0%

is significant; such a difference could be expected to occur by chance not more than twice in 100 random samples.[1] The

[1] The ratio of any difference to its own standard error is called t. With the t value one can enter 'Student's' tables and find the p value. When p is .02 (which signifies that the difference in question might be expected twice in 100 samples drawn at random from a homogeneous population) or less, the difference is regarded as significant. A complete description of this statistical approach is given by Fisher (2).

TABLE II

RESULTS OBTAINED FROM THE CONTROL ANIMALS

The percent scores represent the numbers of times conditioned flexion occurred to the conditioned stimulus. Twenty stimulus-presentations were given during each test-period.

	Group BC				Group LC			
Animal.....	No. 9	No. 10	No. 11	No. 12	No. 13	No. 14	No. 15	No. 16
	Bell + shock to left forepaw				Light + shock to left forepaw			
1	30%	5%	10%	5%	5%	0%	10%	5%
2	40%	30%	35%	30%	0%	0%	10%	5%
3	15%	80%	25%	75%	10%	0%	40%	0%
4	30%	85%	65%	95%	15%	0%	20%	0%
5	30%	85%	95%	85%	80%	0%	5%	20%
6	25%	40%	95%	95%	60%	0%	45%	20%
7	90%	90%	90%	85%	85%	0%	75%	25%
8	90%	85%	95%	95%	90%	0%	80%	35%
9	85%	90%	85%	95%	75%	70%	85%	65%
10	95%	90%	100%	100%	80%	75%	85%	85%
11	95%	95%			85%	90%	95%	75%
12	90%	100%			100%	80%	95%	75%
13	100%					90%	100%	95%
14						90%		90%
15						85%		100%
16						85%		
17						95%		
18						95%		
19						95%		
20						100%		
Critical Tests								
	Light alone				Bell alone			
1	0%	0%	0%	0%	0%	0%	20% 0%	0%

experimental animals required a total of 27 test-periods (a mean of 3.75 test-periods; S.E. = 0.85 test-periods) to reach the zero-response criterion. The control animals required a total of 9 test-periods (a mean of 1.12 test-periods; S.E. = 0.12 test-periods) to reach the same criterion. The mean difference, 2.63 test-periods (S.E. = 0.86 test-periods) in favor of the experimental group, is highly significant. When subjected to Fisher's technique, such a difference could be expected by

chance only thrice or less in 1000 samples drawn at random. Therefore, animals given 200 successive experiences of two discrete sensory stimuli in contiguous combination will respond with forelimb flexion to one of them, when the other one *alone* has been made a conditioned stimulus for forelimb flexion by appropriate training with shock. The differences both in frequency and persistence of response to the stimulus never presented with shock are of statistical significance in favor of the experimental animals.

It is apparent that all of the animals of group LE responded to bell in the critical tests whereas only two of the animals of group BE responded to light. Group LE, then, responded a mean of 6.00 (S.E. = 8.94) times more often and a mean of 1.25 test-periods (S.E. = 1.79 test-periods) longer to bell, than the animals of group BE responded to light. These differences, while not significant, both being expected by chance 25 in 100 times, do, however, point to a greater effect from the bell than from the light. This tentative conclusion becomes strengthened when we find that the animals conditioned to light (both experimental and control) required a mean of 60 trials more to attain the 100 percent conditioned response-criterion than did the animals conditioned to bell. This difference is statistically significant and gives evidence of the greater strength of the bell. Furthermore, the bell is rung against a background of silence, while the light, although emitted by an unfrosted 150 watt bulb, appears upon a background of high luminous intensity. The sound differential is thus considerably greater than is the differential in visual brightness. Therefore it is assumed, that because of the disparity in the intensity of the two stimuli, the measures of sensory preconditioning (frequency and persistence of flexion-response) are greater when a test is made with the stronger of the paired stimuli (bell) than with the weaker one (light).

The experience of 200 trials of bell and light did not effect the rate of conditioned response acquisition. The animals of group BE required a mean of 15 more trials to reach 100 percent conditioning than did those of group BC. The ani-

mals of group LE required a mean of 15 trials less to reach the 100 percent response-criterion than their controls, group LC. In both cases, the differences are of chance expectancy.

Discussion

That the flexion-response is elicited by the stimulus not associated with shock must be due to an association formed when the bell and the light were presented contiguously, prior to the conditioning of flexion to one of them. For the control animals, which were given identical treatment except for the 200 trials of bell and light in combination, did not respond, or responded very infrequently to the stimulus which had never been presented with shock. Some 'bond' must have been formed between the bell and the light, and some 'trace' of this bond retained. Such a 'bond' and its 'trace' can only be inferred, for the phenomenon observed experimentally is essentially this. Flexion becomes conditioned to a stimulus which has never been associated with shock, the normal procedure which produces this conditioning. This phenomenon is dependent upon two preceding events. First, this stimulus must be presented contiguously with a second stimulus for a number of times. Then, the second stimulus must be combined with shock until forelimb flexion to it has been firmly established. When these two conditions have been fulfilled, the first stimulus when given alone will evoke flexion. Therefore, we infer a sensory conditioning between the bell and light from the behavior observed experimentally, sensory preconditioning.[2]

It is difficult to make a conjecture as to the nature of the relationship between the bell and the light. It is extremely unlikely that an artificial synæsthesia was produced, for Kelley (3) was unable to develop an artificial chromæsthesia in humans by presenting tones of different pitches simultaneously with lights of various colors. Even under the most favorable conditions for the appearance of chromæsthesia

[2] Cason (1) has used the term sensory conditioning to describe the effect of conditioning one of two sensory stimuli upon the judgment of their relative intensities.

(startle, fatigue, and mescal intoxication), no evidence of a conditioned chromæsthesia was procured. Suffice it to say that since one of the stimuli will elicit flexion, when the other alone has been combined with shock, some relationship must have been established between the two stimuli when they were presented together.

Since sensory pre-conditioning is a phenomenon of phasic activity, it is of importance as a form of behavior-modification. It is a basic principle of psychology that organisms are capable of making responses in the light of past experience. Much of this modification in behavior takes place under circumstances which fit the principles of conditioning. It is likely that in the commonplace environment of any organism, the conditions of the experiment on dogs herein reported will be fulfilled in nature. The experience of two contiguous sensory stimuli completely divorced from any phasic activity is frequent to any organism. If one of these stimuli becomes the signal for the response of a given reaction-pattern, the other will then elicit a similar response.

Conclusions

1. Dogs given 200 successive combinations of bell and light will respond with flexion to one of these when the other has been made a conditioned stimulus for flexion by appropriate training with shock.

2. The response to the stimulus not associated with shock is due to the prior association of this stimulus with another stimulus, this other stimulus having been made a conditioned stimulus for flexion; for the control animals, which were not given bell and light in combination, did not respond, or responded very infrequently to the stimulus never presented with shock.

3. There is some evidence that the bell is a stronger stimulus than the light, and therefore will elicit flexion more often and for a longer time, once bell and light have been given in combination, and flexion then conditioned to light.

4. The 200 trials of bell and light given in combination

before conditioning did not effect the rate of conditioned response acquisition.

(Manuscript received March 22, 1939)

References

1. Cason, H., Sensory conditioning, *J. exper. Psychol.*, 1936, **19**, 572–591.
2. Fisher, R. A., Application of 'Student's' distribution, *Metron*, 1925, **5**, 90–104.
3. Kelley, E. L., An experimental attempt to produce artificial chromaesthesia by the technique of the conditioned response, *J. exper. Psychol.*, 1934, **17**, 315–341.
4. Prokofiev, G. and Zeliony, G., Des modes d'associations cerebrales chez l'homme et chex les animaux, *J. de Psychol.*, 1926, **23**, 1020–1028.
5. Shipley, W. C., An apparent transfer of conditioning, *J. gen. Psychol.*, 1933, **8**, 382–391.
6. Shipley, W. C., Indirect conditioning, *J. gen. Psychol.*, 1935, **12**, 337–357.

THE INFLUENCE OF ATTITUDE ON THE CONDITIONED EYELID RESPONSE[1]

BY DAVID A. GRANT

University of Wisconsin

Introduction

There has been some tendency lately to explain certain seeming inconsistencies and variations in the results of conditioning experiments in terms of variations in sets or attitudes, but there has been little work testing the relative effects of different attitudes on the conditioning process. Guthrie (**11**) introduced the notion of set to explain experimental extinction and spontaneous recovery. Razran's theory of types of conditioning in man (**27**) was based on different attitudes toward the linking of the new stimulus to the old response.

Loucks (**18**) and Hilgard and Allen (**14**) found that the unconditioned stimulus must do more than force a response in order that conditioning may occur. They concluded that secondary postural or attitudinal adjustments were necessary for conditioning to take place. Razran (**28**) reported that his subjects exhibited unusually regular salivary conditioning when attitudinal and verbal activities were controlled during the reinforcement series. This control was achieved by directing the attention of the subjects to a maze problem the solution of which was external to the actual conditioning experiment. Schilder's (**29**) introspective study of the formation of conditioned hand withdrawal responses suggested that conditioning did not appear until after various states of apprehension and excitement had disappeared. Other data gathered from various sources have been summarized by Mowrer (**20**) who emphasizes the importance of set and attitude in the formation of conditioned responses.

The specific problem undertaken in this study was to as-

[1] This experiment was carried out in the psychological laboratory of the State University of Iowa.

certain the effects on the conditioning process of two attitudes or sets toward the conditioned stimulus. These attitudes were induced by instructions given to the subject at the beginning of the experiment. One group of subjects was instructed to maintain an expectant attentive set toward the conditioned stimulus especially at the moment of appearance. In order to see that this attitude was consistently maintained from trial to trial, the subject was instructed to make a tapping response with his finger when the stimulus appeared. A second group of subjects was instructed to fixate the source of the conditioned stimulus but in other respects to adopt a passive attitude toward the stimulus.

Miller and Cole (**19**) made a study of the conditioned eyelid response using a shock to the cheek as the unconditioned stimulus and light and a tug at the wrist as the conditioned stimulus pattern. In the procedure used, the tug at the wrist was preceded by the light stimulus and followed by the shock to the cheek. There was more rapid and regular conditioning of the eyelid response in a group of subjects who made a finger tapping response to the tug at the wrist than in a group which did not make the finger response. Miller and Cole seemed to emphasize the influence of the overt tapping activity on the formation of conditioned eyelid responses rather than the influence of the more general attentive attitudes of the subjects in the tapping group.

The particular response used in the present study was the eyelid response, the conditioning of which was originally reported by Cason (**4**). Later studies of the conditioned eyelid response in human subjects have been carried out by Cason (**6**), Dodge (**9**), Switzer (**31**), Hilgard (**12**; **13**), Cohen, Hilgard, and Wendt (**8**), Shipley (**30**), Cason and Katcher (**7**), Bernstein (**1**), Neet (**21**), Miller and Cole (**19**), Hilgard and Campbell (**15**), Porter (**24**; **25**; **26**), and Campbell (**2**). We used the type of apparatus and photographic recording device which were developed by Dodge (**10**), and which have been used in similar experiments by Hilgard and others. The relatively high degree of specificity of the lid reflex makes it an excellent response to use in an investigation of this kind since

widespread postural adjustments of the body in different attitudes are not likely to affect the eyelid musculature directly.

Procedure

General Procedure.—A modification of the Dodge pendulum photochronograph was used to control the temporal relations of the stimuli and to record the movements of the eyelid. This apparatus is described by Dodge (**10**), and its use in experiments with the eyelid response has been reported by Hilgard (**12**), Peak (**23**), and others. The original or unconditioned stimulus for the right eye was provided by a puff of air released sharply at a distance of 20 mm from the subject's right eye. The air pressure before release was 150 mm of mercury, except in the case of two subjects to be mentioned later, and the puff caused complete closure in all except one subject. The substitute or conditioned stimulus was provided by a change from 0 to 1.5 foot-candles in the illumination of a circular milk glass window 75 mm in diameter placed 60 cm in front of the subject's eyes. The puff of air was delivered approximately .45 sec. after the initiation of the light stimulus, and the light stimulus remained on until after the eyelid reaction was completed.

The subjects were divided into two experimental groups on the basis of the following instructions which were read and in most cases discussed with the subjects.

Active group. "Adopt a very expectant and attentive attitude toward the milk glass disk and the light stimulus. Wait for it! Be ready for it! When the light appears you are to make a quick response to it by tapping with the index finger of your right hand. Try to relax your body otherwise, but be sure to watch for the light. You will be told to wink before each trial which you are to do immediately. This will serve as a ready signal. Please report any disturbances, distractions, or failure to follow these instructions. Are there any questions?"

Passive group. "Relax and adopt a completely passive and indifferent attitude toward the experiment. Fixate the milk glass disk but do not attend to it particularly. Try to let your eyelid remain passive and relaxed. You will be told to wink before each trial, which you are to do immediately. This will serve as a ready signal. Please report any disturbances, distractions, or observations leading to artifacts. Are there any questions?"

Subjects.—The subjects were 16 students and staff members in the psychology department at the State University of Iowa, and all were at least familiar with the principle of the conditioned response. In the active group were four men and three women, and in the passive group were six men and three women. Data from two additional subjects were not analyzed because the nature of the original response to light was such that the later existence of a conditioned response would have been almost impossible to detect.

Procedure with an Individual Subject.—With very few exceptions, each subject came at the same time on each of the four consecutive days of the experiment. The first and also the last period varied from 15 to 45 minutes, while the second and third periods lasted from 45 to 60 minutes. On the first day, the subject was given instructions, and an artificial eyelash was attached to the right eyelid. Records were made of the unconditioned responses to light alone and to the air puff. We also measured the voluntary reaction time of winking to the light stimulus. No paired stimulations of light and puff were presented on the first day.

On the second and third days, the instructions were repeated, and the artificial eyelash was glued to the eyelid. The order of stimulation was as follows.

5 presentations of light alone
50 presentations of light followed by the air puff given .45 sec. after the beginning of the light
5 presentations of light alone.

This sequence was used with human subjects by Hilgard and Campbell (15), but they recorded the response of the non-puffed eye. The time between trials varied from about 30 to 60 sec. Rest periods were inserted every tenth trial after the first ten.

On the fourth day, the subject was stimulated with the light alone for 20 trials; but if the response failed to appear or if it disappeared quickly, fewer trials were given.

A noise screen with a loudness level of 75 db. masked the operating noises of the stimulus switches and pendulum. We wish to acknowledge the aid of Dr. Don Lewis of the State University of Iowa for estimating this loudness level. During each day's session, we tested for the presence of incidental stimuli by going through the whole procedure of setting and releasing the pendulum and switches without presenting the conditioned and the unconditioned stimuli. No evidence for the conditioning of the eyelid response to incidental stimuli was found. Visual stimuli were minimized since the experiment was conducted in semi-darkness. We used a procedure similar to Campbell (16, 567) in presenting the ready signal 'Wink!' to keep the incidental winking from complicating the results.

Photographs of records similar to the ones obtained in this study have been reproduced in several articles (12; 15; 17). The records of the stimuli and the movements of the eyelid appear on a graph-like background formed by concentric curved abscissæ and straight ordinates, the ordinates indicating .01 sec. intervals. In these records time is measured from left to right, and the eyelid reaction appears as an upward curve in the line which represents the original position of the eyelid. Since the artificial eyelash serves as a lever, the motion at the eyelid is magnified four to five times on the record, and this error is somewhat greater in responses of larger amplitude. If the radian measure of the angular motion of the eyelid be designated by θ, this error is roughly proportional to $\tan \theta - \theta$.

Results

The initial responses to the light-puff stimulus complex were much more diffuse than the responses later in the series. Startle reactions at first ranged from general body movement to slight movements of the eyelid. In the case of one subject it was necessary to increase the air pressure and with another subject it was necessary to decrease the air pressure in order to secure an adequate unconditioned response of the eyelid alone. Slight irritation of the cornea was reported by a few subjects, but this irritation generally disappeared quickly. The initial overt responses decreased in extent as the experiment progressed.

All the subjects showed at least ten anticipatory eyelid closures during the 100 reinforcing trials on the second and third days. An anticipatory wink is one in which the conditioned closure occurs before the air-puff. The phenomenon of anticipation is quite common in eyelid conditioning when certain time intervals are used (**17**; 433–434).

Stable conditioned eyelid responses were obtained from all seven subjects of the active group, and in three out of nine subjects in the passive group. Hilgard and Campbell (**15**) report that under somewhat similar conditions, over half of their 63 subjects showed stable conditioned responses.

Verbal Reports of the Subjects.—With the exception of one subject in the passive group, our subjects reported that their eyelid responses were involuntary. In the case of the subject mentioned above, for whom the pressure of the air for the unconditioned stimulus had to be lowered, irritation of the eye from the puff caused him to wink voluntarily to avoid further pain. Lowering the air pressure enabled him to relax, and he no longer wished voluntarily to avoid the puff stimulus.

As the conditioning progressed, several subjects reported that each time they heard the ready signal 'wink!,' they felt a mounting compulsion to close the eyelid. This compulsion became stronger when the light appeared and was so great that they were forced to close their eyes for a moment. This increasing compulsion to close the eyelid is the same kind of process of expectancy that was described and discussed by Mowrer (**20**). These reports of compulsion and expectancy are similar to those given by Schilder's subjects (**29**). In his introspective study of the process of conditioning the hand withdrawal response, Schilder stated that at least two of his subjects were not conditioned in the usual sense of the term, but that in these two cases the psychological processes of expectancy and desire to withdraw the hand were conditioned.

The relative degree to which this expectancy is verbalized should be an important factor in determining the kind and direction of the resulting behavior. Verbal activities of the subjects have caused considerable confusion in the interpretation of the results of conditioned response experiments.

Cason (**5**) has discussed some of the general results of uncontrolled verbal activities in human conditioning, and Razran (**28**) has shown that salivary conditioning proceeds more regularly when the verbal activities of the subject are directed to an activity outside of the conditioned response situation.

The reports of subjects in the present study showed that the psychological activities involved in the experimental situation were verbalized in various degrees. One subject reported that he felt as if he had forgotten something, and then there seemed to be a sudden command to wink. Another subject showing less verbalization reported a vague expectation of something, accompanied by a feeling that his eye was slowly closing. A systematic study of verbalization and resulting behavior during conditioning could do much to clear up the problem of controlling verbal processes with human subjects.

Initial Appearance of Conditioned Responses.—One of the great advantages of Dodge's photographic technique lies in the fact that the end result of each trial is a graphic record of the complete response of the eyelid and the temporal relations of this response to the occurrence of the conditioned and unconditioned stimuli. The anticipatory character of the conditioned eyelid response separates it from the unconditioned response to the puff, and in most subjects this anticipatory response is not likely to be confused with the slight unconditioned wink which is the natural response to light. In the present study, this unconditioned response to light occurred on the average about .15 sec. before the conditioned wink. As has been pointed out before (**12**), this permits investigation of the conditioned response without inserting test trials which might have an undesired effect on the conditioning process.

On the first day of paired stimulation (second day of the experiment), the first appearance of the anticipatory response occurred on the average on trial number 9 of the conditioning series (trial number 14 for the day) for active subjects, but not until trial number 17 for the passive subjects. The second day of reinforcement (third day of the experiment) produced a smaller difference, the initial appearance of the

conditioned response in the active group occurring on the average on trial 3 as compared with trial 5 for the passive group. In Campbell and Hilgard's experiment (3), the *fifth* anticipatory response occurred on the average on trial number 25. The attitudes of their subjects would correspond best to the attitudes of our passive group whose *first* response occurred on trial 17 of the first day.

Frequency of Conditioned Responses.—The conditioning series of 50 trials on day two and 50 trials on day three was divided into ten successive blocks of ten trials each. The proportion of conditioned responses occurring in each successive block for the active and passive groups is presented in the left-hand portion of Table 1. The active group shows a greater percent conditioned responses in each block than does the passive group. These same data are plotted graphically in the upper part of Fig. 1. The curves of the active group are

TABLE 1

CHARACTERISTICS OF THE CONDITIONED RESPONSES DURING REINFORCEMENT TRIALS ON THE TWO DAYS OF CONDITIONING, AVERAGED FOR SUCCESSIVE TEN-TRIAL BLOCKS

	Trials	Frequency of Occurrence (Percent)		Amplitude (Recorded in mm)		Latency (Hundredths of a sec.)	
		Active Group N=7	Passive Group N=9	Active Group N=7	Passive Group N=9	Active Group N=7	Passive Group N=9
Day 2	6–15	18.4	8.0	6.2	2.1	28.9	26.0
	16–25	38.8	29.7	13.4	8.8	31.6	28.8
	26–35	40.0	29.7	15.7	8.9	30.6	27.2
	36–45	49.0	41.0	19.5	13.0	29.0	33.2
	46–55	68.5	45.7	23.2	12.6	27.2	29.8
Day 3	6–15	64.5	45.7	20.0	13.7	27.4	29.5
	16–25	86.0	67.8	29.8	23.0	26.7	29.0
	26–35	93.0	72.0	30.4	22.2	25.9	27.7
	36–45	91.3	80.0	33.3	29.2	26.7	29.7
	46–55	94.0	65.7	31.8	23.9	26.4	28.4

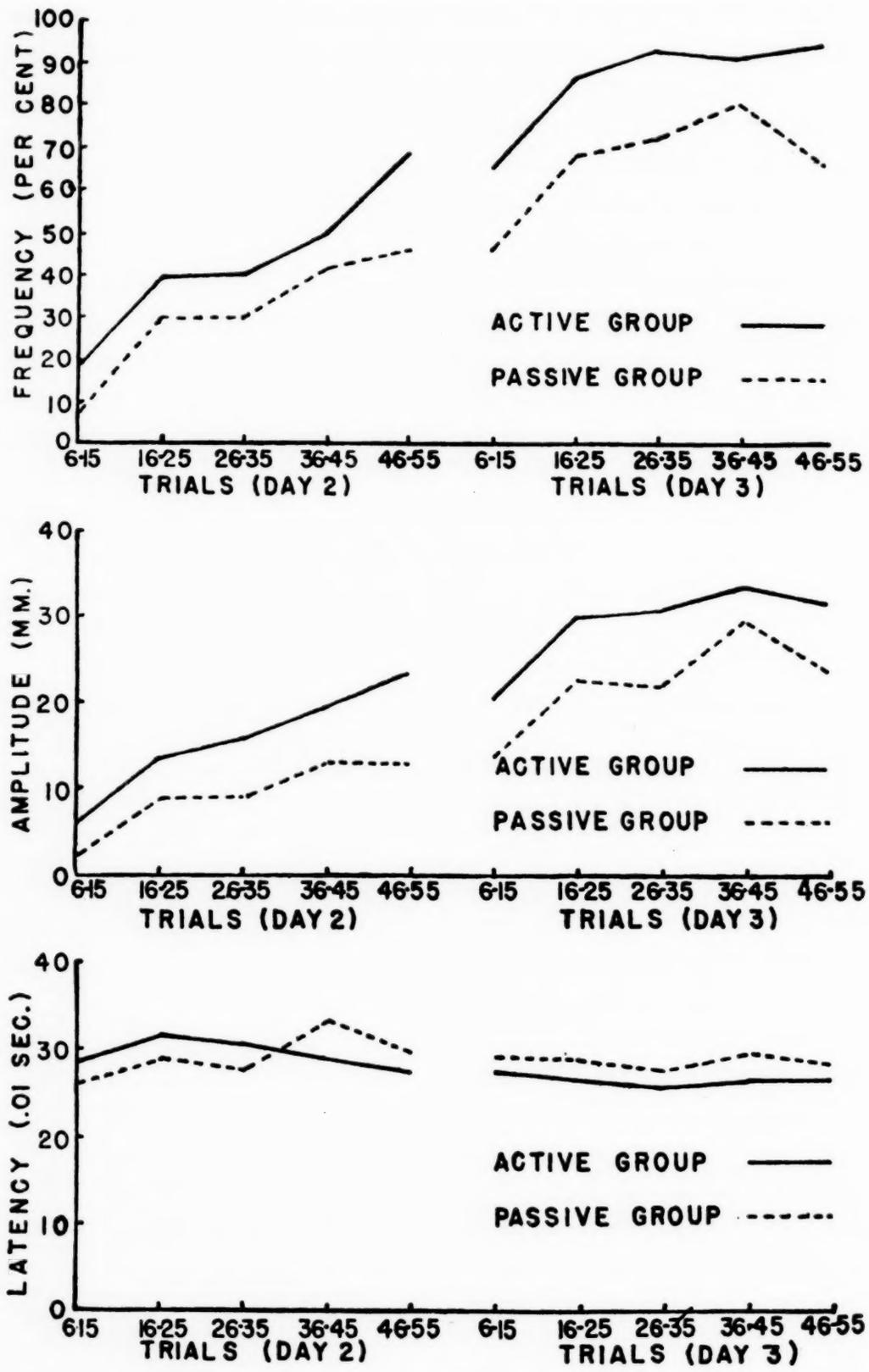

FIG. 1. Characteristics of the conditioned responses averaged for successive ten-trial blocks during the two-day conditioning period.

somewhat more regular than those of the passive group. This result harmonizes with that obtained by Miller and Cole (**19**), and they also found a greater relative frequency of conditioned responses in their active group. The flat sections in the frequency curves of day two between trials 16–25 and 26–35 may have been due to fatigue or some change in attitude, although these were not consistently reported by the subjects. The drop in frequency for the passive group at the end of day three was also quite possibly due to a fatigue or change in attitude.

The average number of anticipatory conditioned responses per subject for the second day of the experiment was 25.2 for the active group and 15.3 for the passive group, and the average number of conditioned responses on the third day was 43.0 for the active group and 33.3 for the passive group. The average number of conditioned responses for each of our passive subjects during the two-day period was 48.5 as compared with 62 reported by Campbell and Hilgard (**3**). The average number of conditioned responses from each of the active subjects was 68.2.

The average number of conditioned responses in the short extinction series of 20 or fewer trials on day four showed that the conditioning in the active group (10.4 conditioned responses per subject) was more stable than the conditioning in the passive group (4.5 conditioned responses per subject).

Amplitude of Conditioned Responses.—Measurements of amplitude are subject to certain errors in recording which have been mentioned above. Other difficulties are occasioned by the fact that the amplitude for complete closure of the eyelid varies from subject to subject, so that an individual record of 25 mm amplitude might mean maximum closure for one subject but only two-thirds closure for another.

As in the case of the frequency measures, the recorded amplitudes have been averaged for each group of subjects for each successive ten-trial block. These averages are presented in the central portion of Table 1. As in the case of the frequency data, results on amplitude in each block also show that there was stronger conditioning in the active group. These data on amplitude are plotted in the central portion of Fig. 1.

As with the frequency curves, the amplitude curves of the active group are somewhat more regular than those of the passive group.

On day two the average amplitude of all the conditioned responses was 15.6 mm for the active subjects and 9.4 mm for the passive subjects. On day three the corresponding averages were 29.0 mm for the active group and 22.4 mm for the passive group. There was a good deal of overlapping between the amplitude data of the groups. The individual subjects showed curves of the same general shape, but the individual curves were more irregular than the average curves.

The average amplitude of all conditioned responses of our passive group was 15.9 mm as compared with 16.6 mm reported by Campbell and Hilgard (**3**).

Latency of Conditioned Responses.—We shall use latency to mean the temporal interval between the beginning of the conditioned stimulus (light) and the beginning of the conditioned response (wink). The latency data are presented in the right-hand portion of Table 1. In averaging the amplitudes for each ten-trial block, absence of the conditioned response was treated as an amplitude of zero; but in the case of latency, we followed the procedure of Hilgard and Campbell (**15**) and included in the averages only trials which showed conditioned responses. The average latencies are presented in graphic form in the lower half of Fig. 1. These curves show that the conditioned responses of the passive group had a lower average latency between trials 6–15 and 26–35 on day two but thereafter the responses of the active group had a lower average latency. The meaning of this irregularity is uncertain, but latency measures do not seem to be as stable as other measures of conditioned responses. Campbell and Hilgard (**3**) reported a reliability coefficient of .51 for latencies as compared with .75 and .66 for amplitudes and frequencies, respectively.

On day two both groups had the same average latency (29.4 hundredths of a sec.), but on day three the average latency of the active group was 26.7 and the average latency of the passive group was 28.8 hundredths of a sec. Since

lower latency is associated with stronger conditioning (**22**; **17**) these data on latencies partially support the other results reported on amplitudes and frequencies in indicating stronger conditioning in the subjects of the active group.

We have not computed statistical measures of the reliability of any of the differences between active and passive groups because of the changing psychological conditions of the experiment, especially as shown by the rise in the curves of frequency and amplitude.

Discussion of Quantitative Results.—All measures of conditioning—initial appearance, stability, frequency, amplitude, and latency—seem to indicate that conditioning was stronger in the group instructed to maintain an active attentive attitude toward the conditioned stimulus, but it would be more accurate to say that conditioning proceeded at a more rapid rate in the active group. Analysis of the records revealed no qualitative differences between the two groups in the process of conditioning,[2] and the verbal reports of the subjects were similar for the two groups as far as the conditioning process itself was concerned.

There is a certain amount of ambiguity as to what variable caused the more rapid conditioning of the active group. Miller and Cole (**19**) claimed that tapping promptly with the finger in response to a tug at the wrist facilitated the formation of conditioned eyelid responses. They also recognized that their subjects had to maintain a relatively high degree of attention to make the prompt finger responses. Razran (**28**) found more regular conditioning of the salivary response when his subjects were busy solving a maze problem. He interpreted the better results in terms of his subject's consistent and controlled attitudes directed away from the conditioning process. There seems to be an increased postural tonus in expectancy and heightened attention, and the attentive subject is more active than a non-attentive subject.

Further research will clear up the problem of whether

[2] Further data on qualitative differences between the two groups of subjects will be presented in a forthcoming paper.

increased tonus or specific overt activities or both cause facilitation of conditioning, but in the meantime it is evident that combinations of bodily activity and heightened attention have considerable influence on the process of conditioning.

Summary and Conclusions

Conditioned eyelid responses were obtained from 16 subjects using light as the conditioned stimulus followed by a puff of air to the eye as the unconditioned stimulus. Movements of the eyelid were recorded by an adaptation of the Dodge photochronograph. Seven subjects formed an active group, instructed to maintain an attentive, expectant attitude toward the conditioned stimulus (light) and to make a tapping response when the light appeared. Nine subjects formed a passive group, instructed to adopt a passive attitude toward the conditioned stimulus.

The principal results of the experiment were as follows.

1. All subjects reported that their conditioned eyelid responses were involuntary.

2. The majority of the subjects reported a mounting expectation of the puff during each trial, and this resulted in an impulse to close the eye.

3. The initial conditioned responses were obtained at an earlier trial on the average with the active group than with the passive group.

4. Conditioned responses occurred more frequently in the active group.

5. Conditioning was more stable in the active group.

6. The average amplitude of the conditioned responses was greater in the active group.

7. The average latency of the conditioned responses was lower in the active group, but this trend was not uniform during the entire reinforcement series.

(Manuscript received March 17, 1939)

References

1. Bernstein, A. L., Temporal factors in the formation of conditioned eyelid reactions in human subjects, *J. gen. Psychol.*, 1934, **10**, 173–197.

2. Campbell, A. A., Interrelations of two measures of conditioning in man, *J. exp. Psychol.*, 1938, **22**, 225–243.
3. Campbell, A. A., and Hilgard, E. R., Individual differences in the ease of conditioning, *J. exp. Psychol.*, 1936, **19**, 562–571.
4. Cason, H., The conditioned eyelid reaction, *J. exp. Psychol.*, 1922, **5**, 153–196.
5. Cason, H., The role of verbal activities in the conditioning of human subjects, *Psychol. Rev.*, 1934, **41**, 563–571.
6. Cason, H., Backward conditioned eyelid reaction, *J. exp. Psychol.*, 1935, **18**, 599–611.
7. Cason, H., and Katcher, N., An attempt to condition breathing and eyelid responses to a subliminal electrical-stimulus, *J. exp. Psychol.*, 1933, **16**, 831–842.
8. Cohen, L. H., Hilgard, E. R., and Wendt, G. R., Sensitivity to light in a case of hysterical blindness studied by reinforcement-inhibition and conditioning methods, *Yale J. Biol. & Med.*, 1933, **6**, 61–67.
9. Dodge, R., Problems of human variability, *Science*, 1924, **59**, 263–270.
10. Dodge, R., A pendulum photochronograph, *J. exp. Psychol.*, 1926, **9**, 155–161.
11. Guthrie, E. R., *The Psychology of Learning*, 1935.
12. Hilgard, E. R., Conditioned eyelid reactions to a light stimulus based on the reflex wink to sound, *Psychol. Monogr.*, 1931, **41**, no. 184, pp. 50.
13. Hilgard, E. R., Modifications of reflexes and conditioned reactions, *J. gen. Psychol.*, 1933, **9**, 210–215.
14. Hilgard, E. R., and Allen, M. K., An attempt to condition finger reactions based on motor point stimulation, *J. gen. Psychol.*, 1938, **18**, 203–207.
15. Hilgard, E. R., and Campbell, A. A., The course of acquisition and retention of conditioned eyelid responses in man, *J. exp. Psychol.*, 1936, **19**, 227–247.
16. Hilgard, E. R., Campbell, A. A., and Sears, W. N., Conditioned discrimination the development of discrimination with and without verbal report, *Amer. J. Psychol.*, 1937, **49**, 564–580.
17. Hull, C. L., Learning: II. The factor of the conditioned reflex. In *A Handbook of General Experimental Psychology*, Ed. by Murchison, C., 1934.
18. Loucks, R. B., The experimental delimitation of neural structures essential for learning: the attempt to condition striped muscle responses with faradization of the sigmoid gyri, *J. Psychol.*, 1935, **1**, 5–44.
19. Miller, J., and Cole, L. E., The influence of 'voluntary' reaction upon the development and the extinction of the conditioned eyelid reaction, *J. genet. Psychol.*, 1936, **48**, 405–440.
20. Mowrer, O. H., Preparatory set (expectation) a determinant in motivation and learning, *Psychol. Rev.*, 1938, **45**, 62–91.
21. Neet, C. C., A comparison of verbal, manual, and conditioned response methods in the determination of auditory intensity thresholds, *J. exp. Psychol.*, 1936, **19**, 401–416.
22. Pavlov, I. P., *Conditioned Reflexes: an Investigation of the Physiological Activity of the Cerebral Cortex*, Tr. by Anrep, G. V., 1927.
23. Peak, H., Modification of the lid-reflex by voluntary induced sets, *Psychol. Monogr.*, 1931, **42**, no. 188, pp. 68.
24. Porter, J. M., The modification of the conditioned eyelid responses by successive series of non-reinforced elicitations, *J. gen. Psychol.*, 1938, **19**, 307–323.
25. Porter, J. M., Backward conditioning of the eyelid response, *J. exp. Psychol*, 1938, **23**, 403–410.

26. PORTER, J. M., Experimental extinction as a function of the interval between successive non-reinforced elicitations, *J. gen. Psychol.*, 1939, **20**, 109–134.
27. RAZRAN, G. H. S., Conditioned responses: an experimental study and a theoretical analysis, *Arch. Psychol.*, 1935, no. 191.
28. RAZRAN, G. H. S., Attitudinal control of human conditioning, *J. Psychol.*, 1936, 2, 327–337.
29. SCHILDER, P., Conditioned reflexes, *Arch. Neurol. & Psychiat.*, 1929, **22**, 425–443
30. SHIPLEY, W. C., An apparent transfer of conditioning, *J. gen. Psychol.*, 1933, 8, 382–390.
31. SWITZER, ST. C. A., Backward conditioning of the lid reflex, *J. exp. Psychol.*, 1930, 13, 76–97.

PUPIL DILATATION AND DARK ADAPTATION [1]

BY ROBERT H. BROWN AND HOWARD E. PAGE

Clark University

One finds occasional references in the literature to a correspondence between the pupillomotor and sensory effects of light stimulation. If variation in pupil size were related to the variation in sensitivity which accompanies such changes in the illumination as wave-length, time, and intensity, we might conclude that the iris musculature is functionally dependent upon the same processes which underlie the variation in sensitivity. It would not be surprising if such were the case since the pupil is equipped with its reciprocal action of radial and sphincter fibers to dilate with increased sensitivity and to constrict with decreased sensitivity. Various investigators have obtained luminosity values by measuring the diameter of the pupil when illuminated by light of various wave-lengths (Sachs, **26**; Abelsdorff, **1**; Laurens, **18**). The variation in pupil diameter with wave-length for the human eye is roughly comparable in shape and position for a given level of adaptation to the more conventional luminosity curve obtained by measuring the brightness for various wave-lengths throughout the spectrum. The Purkinje phenomenon is apparent in a shift of the pupil luminosity curve toward the red with increased illumination.

The appearance of the Purkinje phenomenon in the relative efficiency of colored lights upon pupillary diameter in the dark- and light-adapted eye seems to conflict with evidence that variation in pupil size with illumination is almost exclusively controlled by cones in the macular region of the retina. Hess has found that lighting of the macula yields the best pupilloconstriction of any part of the retina and has restricted pupil excitation to a small central region of the retina.

[1] Expenses of this study were met in part by a grant-in-aid from the National Research Council to one of the authors (R. H. B.).

He argued that where pupil constriction occurs to peripheral stimulation it is due to refracted and diffused light reaching the macula. In part Hess's evidence for the influence of dispersed light consisted in the demonstration (**16**) that projection of light upon the blind-spot evokes a distinct pupil constriction. He observed no change in the pupil when the image of the light source was moved outward from the blind-spot.

Vervoort (**28**) found that pupil diameter was uninfluenced by an increase in area of stimulation between the limits of 4.3° and 8.6° visual angle if the quantity of light remained constant, but that pupil diameter decreased from 4.5 mm to 3 mm with a fourfold increase of area without a change of light intensity. As Abelsdorff and Feilchenfeld (**2**) pointed out, these limits do not preclude macular control of pupil size. Abelsdorff and Feilchenfeld found that for both monocular and binocular illumination of the light-adapted eye increased area beyond 3.6° visual angle had little influence upon the relation between pupil constriction and intensity. When the dark-adapted eye was stimulated by a large area, the center of which was black, however, the pupil reflex was evoked. Pupilloconstrictions occurred to stimulation by a circular area with a diameter of 11.3° visual angle where the circular center area of 6.8° diameter was covered with black paper, and also to stimulation by an area whose diameter subtended 45° visual angle but whose center area, 38.7° in diameter, was covered. More recent work by Ferree, Rand, and Harris (**8**) and Luckiesh and Moss (**20**) has shown that pupil size under constant illumination is influenced by stimulus area greater than 20° visual angle, but the effect of increased area outside the macula is negligible.

As Engelking (**6**) has pointed out, no definite conclusion can be drawn from these latter experiments that extramacular regions may initiate pupil constriction. Light entering the eye may be scattered on the retina by reflection at the surfaces of the lens, cornea, and retina. When light falls upon the periphery, it is possible that excitation may be transmitted through the retinal layers from the receiving receptors to the

ganglion cells of the fovea. Finally there may be leakage of excitation from nerve fiber to nerve fiber in the optic nerve. These possibilities would all agree with the relative inexcitability of pupil constriction to peripheral stimulation.

Engelking determined the minimal intensity required to evoke a pupil constriction to 1″ exposure of a large field after 20′ dark adaptation. He observed two kinds of reactions, an abortive reaction with threshold at .001 m.c. and a complete reaction with threshold at .025 or .04 m.c. He pointed out that the threshold value for definite pupil constriction corresponds closely to the lowest foveal threshold intensity. The absence of pupil constriction or small decrease (.2 mm or less) in pupil width to stimulation by intensities less than .01 m.c. would seem to indicate that constriction is almost entirely controlled by the fovea. Where there is some small constriction present at intensities below the foveal threshold, it is impossible to determine what the result would be with the rods alone without the cofunctioning of the cones. For this reason Engelking suggested that more conclusive evidence might come from cases of total color-blindness.

Engelking (**7**) observed that the pupil in a totally color-blind eye constricted and dilated to the same extent but much more slowly than in the normal or dichromatic eye. Hess (**16**) has observed the absence in total color-blindness of normal fluctuations in pupil size to rapidly alternating intensities of light, but has observed pupil reactions to slowly alternating intensities. Geldard (**11**) has observed the virtual absence of a pupil reflex in a case of total color-blindness. "With light having an intensity of 0.074 f.c. at the eye only a very minute and immeasurable twitch of the pupil was observed. For the same conditions the normal response is a reduction of pupil diameter of the order of 1.6–1.8 mm. The maximum response obtained under any of a wide range of light intensities used was but 0.2 mm. There is thus little doubt that pupil responses play no important rôle in the behavior of this case."

Geldard's observation, together with the above evidence, would seem to indicate that to a very large extent pupil

constriction or dilatation to changed illumination is initiated in the macula where the cones predominate in the normal eye. The Purkinje effect in terms of the Duplicity Theory would then result from transmission of excitation from the rods within or outside the macula to optic nerve fibers leading off from cones and controlling pupil constriction.

The constriction and dilatation of the pupil bear certain superficial resemblances to foveal light- and dark-adaptation. Foveal dark-adaptation is extremely rapid during the first few seconds, so that most of the adaptation takes place during the first 30″ and the process is virtually complete after 10′ (Hecht, **13**). Pupil dilatation as measured by Reeves (**25**) is likewise precipitous at first, with much of it completed in 30″ and the maximal diameter practically attained in 5′, although one observer (Garten, **9**) has noticed continued increase in pupil size for periods ranging from 1 hr. to 8 hrs. Reeves also found that in constricting to a stimulus area with a brightness of 100 ml. after dark-adaptation the average pupil reaches its minimal size in less than 5″, with a greater part of the constriction occurring within the first 2″. Measurements of light-adaptation, where they have been made, have necessarily been indirect, since the process is too rapid for accurate measurement. Lohmann (**19**) determined the course of light-adaptation by measuring the threshold intensity 10″ after cessation of light-adaptation for various intervals. He found rapid decrease in sensitivity with time of light-adaptation. In the case of diffusely reflected daylight, the process was virtually complete within 20″. Geldard (**10**) has presented light to one region of the eye and measured its apparent brightness after exposures of various intervals by comparison with another light variable in intensity and falling in a neighboring region. During exposure to 13,805 m.c., the comparison intensity decreased 69 percent in the first 5″ and 91 percent in the first 10″. The process was virtually complete within 30″. For binocular data where interaction in one retina was avoided, during exposure to 13,805 m.c. the average comparison intensity decreased 56 percent in 5″, 64 percent in 10″, and the adaptation, as measured by the comparison light, was virtually complete in 30″.

No exact comparison can be made of pupillary constriction values and light-adaptation data. It may be pointed out, however, that the decrease in pupil area parallels the increase in visual acuity, critical fusion frequency, retinal potential, and optic-nerve discharge, with increase in intensity of illumination. As indicated by Engelking (**6**), the curve relating maximal pupil constriction and intensity has the same shape and position as curves for other visual functions. Like so many visual functions, the relation of pupil area to logarithm-of-the-intensity of illumination is sigmoid with an approximately linear relationship from .01 ml to 100 ml.

The relation between the course of pupillary dilatation and foveal dark-adaptation is indicated more precisely in Fig. 1, where the pupil-area in square millimeters and the logarithm-of-the-threshold-intensity in millilamberts are plotted against the logarithm-of-time-in-the-dark in seconds. The squares represent average values obtained by Reeves (**25**) from 7 observers whose pupils were photographed by light from a bank of lamps which were turned on after given intervals of darkness following adaptation to a square field subtending a visual angle of about 80° (**24**) and with a brightness of 100 ml. The triangles represent the average values of data obtained by Crawford (**5**) from 12 subjects. Crawford used infra-red radiation which permitted him to photograph the pupil at intervals during dark-adaptation following exposure to the illumination from an opal lamp subtending a visual angle of 10° and with a brightness of 9000 ml. The circles represent the log intensity values (uncorrected for pupil area) determined by Hecht (**13**) for foveal dark-adaptation after 5′ adaptation to a field with a brightness of 90 ml. The value assigned at zero time (beginning of dark-adaptation) is the one Hecht obtained by extrapolation. The curve is one of best fit.

The three sets of data are plotted in Fig. 1 so that the values for pupil area and threshold intensity coincide at the very beginning of dark-adaptation, shown by the break in the curve, and at log $t = 2.26$. Since the average pupil area for Reeves's data increases more rapidly after 1″ and reaches a

larger final value than that for Crawford's data, the scale of pupil area on the ordinate refers only to Crawford's data. Reeves's measurements have been multiplied by a constant factor to indicate their similarity to Crawford's values. The pupil areas obtained by Reeves and Crawford are 6.4 and 6.7 sq. mm., respectively, at the beginning of dark-adaptation, and 44.6 and 39.8 sq. mm. at log t = 2.26. The more rapid increase in pupil area and greater maximal pupil opening observed by Reeves than by Crawford may be due simply to

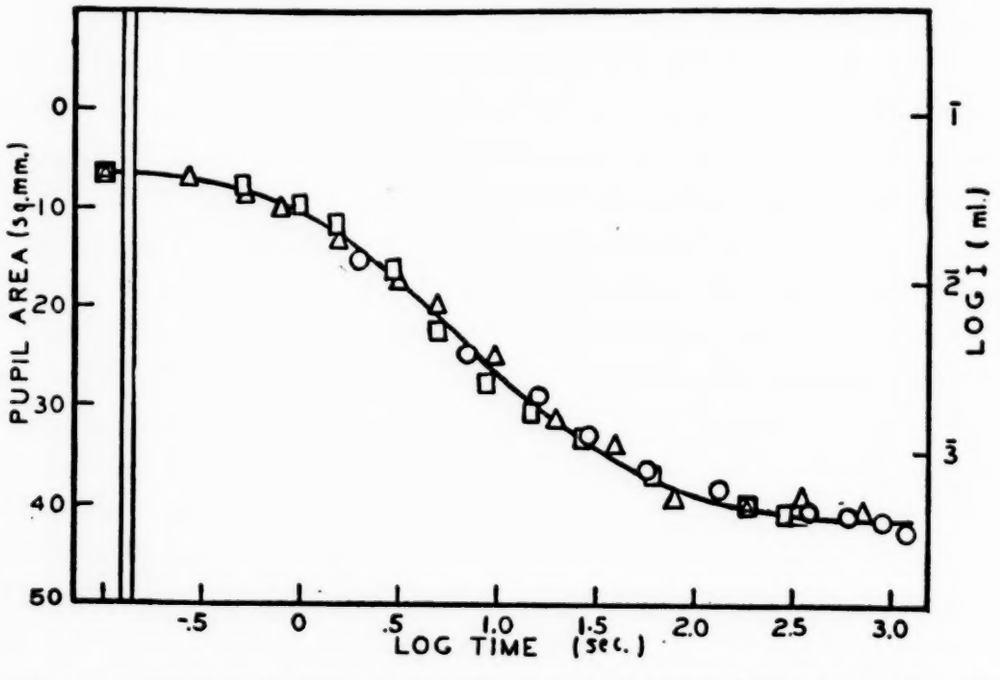

FIG. 1. Relation between pupil dilatation and foveal dark-adaptation. The points represent average values obtained by Reeves (squares) and Crawford (triangles) for pupil area during dark-adaptation and by Hecht (circles) for foveal dark-adaptation. The curve is one of best fit.

differences in their subjects or to differences in their experimental conditions such as brightness and area of the preadapting field.

It may be noted that in general the course of pupil dilatation corresponds to the increase in foveal sensitivity during dark-adaptation. The pupil area varies with duration of darkness in approximately the same manner as log $1/I$. The most systematic deviation occurs in Crawford's data for values greater than log t = 1.9. Crawford's measurements

show virtually no change in pupil area after 80″. This may be due to the dim illumination which was present during Crawford's observations. In view of the general correspondence between variation of pupil area and of log $1/I$ with time of foveal dark-adaptation, we have thought it worth while to obtain rather extensive measurements of pupil dilatation from two subjects. Previous studies have indicated great individual differences in the course of dilatation, some of which may be attributed to individual variations in maximal and minimal size of pupil. The following features of our experiment may be emphasized:

1. The experiment is conducted in a completely dark room free of any extraneous sound which might cause sympathetic dilatation.

2. After 10′ dark-adaptation the observer is light adapted for 2′ to a large stimulus area with a brightness of 1200 ml.

3. At predetermined intervals during dark-adaptation following the 2′ light-adaptation, the pupil of the right eye is photographed with no visible illumination by means of infra-red radiation and film. As early as 1897 Garten (**9**) photographed the pupil with little or no visible illumination by means of ultra-violet radiation. Ultra-violet radiation has the disadvantage that it may act directly on the sphincter muscle (Steinach, **27**; Hertel, **15**). More recently Crawford (**5**), as noted above, and Gullberg, Olmstead, and Wagman (**12**) have determined the course of pupillary dilatation in the human and rabbit eyes by use of infra-red photography which permits a series of photographs of the pupil to be taken after one period of light-adaptation in such a way that the exposures cannot influence the course of pupil dilatation.

4. In all, 15 such series for each of two observers are obtained during the 25′ dark-adaptation following light-adaptation.

5. The observer fixates a small red cross for 10″ before each exposure. Fixation of the red cross, which is kept at a just perceptible brightness by the observer, minimizes variations in pupil size due to accommodation. At the fixation distance we used (75 cm) the pupil diameter is independent

of the distance. Luckiesh and Moss (**20**) have found that pupil size is not altered significantly even by large increases in fixational distance when the latter is 60 cm or more.

Apparatus and Procedure

The apparatus consists essentially of an optical system for binocular light-adaptation, a source of infra-red rays, a Leica camera for photographing the pupil, and timing devices. The stimulus area for light-adaptation, a flash-glass screen, is a square 6 cm on a side and subtends a visual angle of 30°. Light from a 1000-watt Mazda lamp, run on 110-v. direct current, is focused by a plano-convex lens on a diaphragm shutter, used in controlling the period of light-adaptation, after passing through a quartz cell filled with glycerine. When the shutter is open, light, after passing through the desired filter, is reflected by a plane-surface, plate-glass mirror to the stimulus area. Under the conditions of this experiment, the adaptation screen had a brightness of 1200 ml., as calibrated by a Schmidt and Hensch photometer and an optical bench. The mirror and screen used in adapting the observer are placed on a track such that after light-adaptation they can be moved to one side and exposures of the pupil made.

The source of infra-red rays consists of two 1000-watt Mazda lamps inclosed in asbestos-lined, light-tight boxes, one well on each side of the right eye. These lamps are run on alternating current. Reflectors are placed behind the lamps to increase the intensity of radiation. When diaphragm shutters at the front of each box are open, the radiation is centered on the subject's right eye after passing through Wratten infra-red filters No. 87. The complete transmission chart of these filters is given in the booklet, 'Wratten Light Filters,' published by Eastman Kodak Company. This filter transmits no light at a wave-length less than 740 mμ, one-seventh of the radiation at a wave-length of 780 mμ, and one-half at a wave-length of 800 mμ. The infra-red radiation is exposed for 0.1″. Since the radiation transmitted by the filters enters the eye at approximately a 45° angle to the line-of-sight, the faint red image of the lamp filaments, barely visible upon direct fixation, strikes the periphery of the retina and is invisible, even after 30′ dark-adaptation. Control experiments show that the exposure itself has no effect whatsoever upon pupil size. The shutter in the camera remains open throughout the experiment so that the film is exposed when the shutters in back of the infra-red filters are opened. Eastman 35-mm infra-red film is used.

The photographing equipment consists of a Leica camera, Model C, and an Elmar 50-mm lens. A convex 3-diopters lens is fastened to the regular lens holder. The camera is placed 40.5 cm from the eye with the lens set at infinity and the diaphragm at f: 4.5. At this distance a relatively large image of the pupil is obtained. The camera is fitted with a trigger base plate, which facilitates rapid winding of the film, and is fastened in a rigid position by a steel rod screwed to the base plate of the camera and to the table. The subject's head is held in a fixed position by a head-rest, providing four points of contact: two head supports in contact with his temple, a chin rest, and a small brass plate firmly fastened to the table and coated with wax which the subject bites.

When the subject's head is in position, he looks at a small red cross (with bars 1 cm by 2 mm) located 75 cm from the eye and beneath and back of the camera. The subject controls the brightness of the cross during dark-adaptation by adjusting a rheostat. The fixation point serves a double purpose. It keeps the eye in a fixed

position for photographing and aids in eliminating dilatation or constriction of the pupil to changes in accommodation.

The timing devices consist of a long kymograph, signal magnets, an electrically driven tuning fork with a frequency of 100 d.v./sec, a Jacquet chronometer, and stop watches. The signal magnets are wired to record the time at which light-adaptation ends and the time of each exposure. After 25″ the timing is done with stop watches, which can be read to fifths of a second.

After completion of a series the pupil diameter for each exposure is measured by enlarging the negative on a screen. A Leica enlarger is used for this purpose and the readings are accurate within 0.1 mm of pupil diameter. The measurements are calibrated by means of a millimeter rule taped to the subject's cheek bone immediately beneath his eye. The measurements of pupil diameter are not corrected for refraction by the cornea and aqueous humor and represent the photographed external image of the actual pupil. Such a correction would be small, however, and would involve multiplying the observed pupil diameter by a constant factor (1/1.14 for accommodation to an object at infinity and 1/1.02 for an object at 25 cm, as calculated by Nutting, 23, p. 117).

In all 15 series of exposures are made for each of 2 subjects. After 10′ dark-adaptation each subject adapts to light with a brightness of 1200 ml. for 2′ on 15 different occasions. Approximately 30 exposures of the pupil are made during a 25′ period after each light-adaptation. Exposures are made at 2″ intervals for the first 15″ of dark-adaptation. They are then taken at 5″ intervals until 30″ have elapsed. After 30″ exposures are made at increasing intervals until the series is completed. After each exposure the subject advances the film by pulling a cord attached to the trigger on the base of the camera. There are no distractions present which might influence the size of the pupil and the subject concentrates on the fixation point for each exposure. The experimenter is seated in another room from which he controls the shutters determining light-adaptation and film exposure and records the time when light-adaptation ceases and each exposure is made.

Results

The nature of our results is indicated in Fig. 2 where pupil-area in square millimeters and the logarithm-of-the-threshold-intensity in microphotons, determined after comparable conditions of light-adaptation, are plotted against logarithm-of-time in seconds. The triangles represent the average values for pupil area as determined from 15 series of exposures of the right eye of one observer (R. H. B.) after 2′ adaptation to approximately 1200 ml. The circles represent the threshold intensity values obtained by Hecht, Haig, and Chase (**14**) from one subject (S.H.) by measurements with red light after 2′ adaptation to approximately 1200 ml. The drop in threshold for the 5 points after log $t = 3.0$ represents the appearance of rod adaptation. The values for pupil area and threshold intensity have been plotted in Fig. 2

so that they coincide at log t = .68 and 2.90. It may be noted that the increase in pupil area is approximately proportional to the decrease in logarithm-of-the-threshold-intensity. The increase in pupil size after log t = 1.9, absent in Crawford's experiment, while small, is systematic and corresponds to the continued increase in sensitivity. We may conclude that, on the basis of the comparison of data for 2 individuals, pupil area varies directly as log $1/I$ during dark-adaptation. We have not plotted the data for our second subject nor for the 2 other subjects in Hecht, Haig, and Chase's study, since they are quite comparable to those shown in the figure.

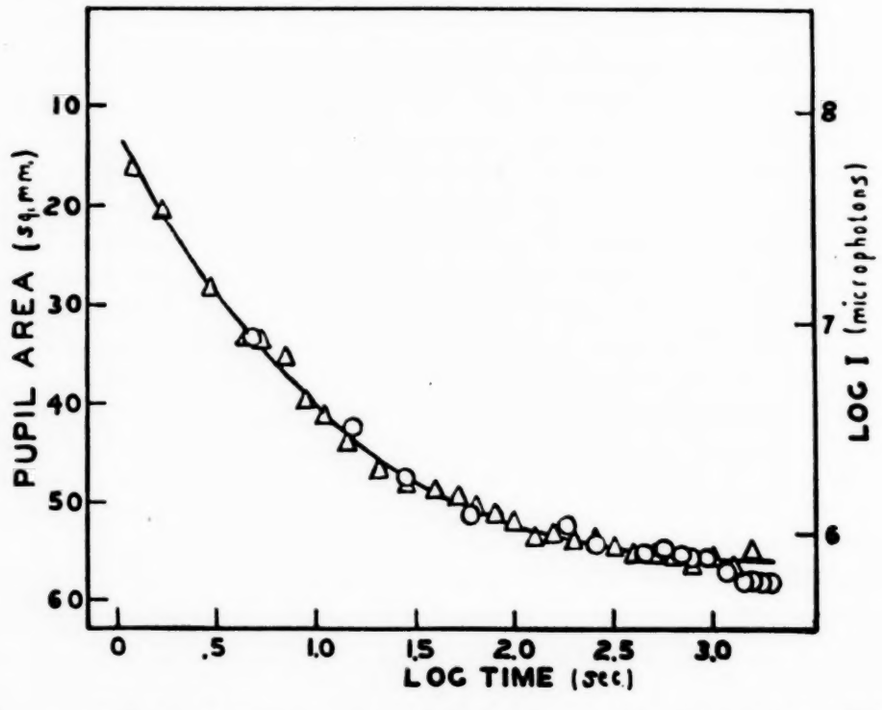

FIG. 2. Relation between pupil dilatation and cone adaptation. Triangles represent average values of pupil area during dark-adaptation; circles, threshold intensities measured with red light by Hecht, Haig, and Chase.

If the relation between pupil size and sensitivity during dark-adaptation has any significance with respect to common processes mediating both, variation in time or intensity of the preadapting light should have the same effect on both. Hecht, Haig, and Chase have found that with increasing intensity of preadaptation the cone dark-adaptation curve is unchanged in form but is shifted along the time axis. Garten

(9) observed a similar displacement in 2 isolated observations of pupil dilatation after weak and intense preadaptation. The discrepancy between Reeves' (25) and Crawford's (5) data is in the direction one would expect on the basis of their preadapting intensities (100 and 9000 ml.). Differences in their experimental conditions (especially fixational distance and preadapting stimulus area) and in their subjects do not permit any exact comparison.

We have made some preliminary measurements which indicate that the displacement along the time axis, with little change in form, of the pupil dilatation curve does occur as the preadapting intensity increases. The same procedure as that described above is followed in obtaining measurements of pupil size during dark-adaptation after greater intensities of the 2′ preadaptation than 1200 ml.

Discussion of Results

The correspondence between pupil dilatation and the increase in cone sensitivity during dark-adaptation is significant for the mechanism controlling pupil dilatation. Pupil area is regulated by the third nerve and cranial autonomic fibers from the ciliary ganglion to the sphincter muscle and by the cervical sympathetic trunk and fibers from the superior cervical sympathetic ganglion to the radial muscle. Apparently dilatation may result from contraction of the radial muscle, or relaxation of the sphincter muscle, or both. Thus, Bain, Irving, and McSwiney (3) observed dilatation to stimulation of the splanchnic nerve after cutting the superior cervical sympathetic trunk in the cat. They concluded that dilatation in this case results from inhibition of the constrictor center in the oculomotor nucleus. Paradoxically, Gullberg, Olmstead, and Wagman (12) found that dilatation in the normal rabbit eye following cessation of light stimulation results from active contraction of the radial muscle under central control and elastic relaxation of the sphincter muscle under local control. Measurements of pupil diameter for the rabbit's eye by infra-red photography showed that a large component of pupil dilatation during dark-adaptation is

eliminated by cutting the cervical sympathetic. The residual dilatation left after sympathectomy is interpreted in terms of elastic recoil of the smooth muscle of the sphincter.

The correspondence between pupil dilatation and increase in cone sensitivity during dark-adaptation may be interpreted in terms of (1) a fortuitous resemblance between two distinct and unrelated processes, (2) an increase in central excitability during dark-adaptation which is manifested in two related processes, or (3) an increase in retinal excitability reflected in two processes under common control. The lack of any evidence for action currents, other than after-discharge, following cessation of stimulation of the retina renders the last interpretation improbable. Although the degree of similarity of the course of pupil dilatation and the increase in cone sensitivity during dark-adaptation argues against the first interpretation, it cannot be eliminated without further study. We are obtaining data on pupil dilatation following various intensities and durations of preadaptation.

The most probable explanation of our results is in terms of changes in the constrictor center of the oculomotor nucleus, or perhaps in the pretectal region, the locus of the first synaptic juncture of the light reflex pathway in the brain (Magoun and Ranson, **22**). When the retina is stimulated by a constant intensity, as in preadaptation, the constrictor center presumably is subject to a constant state of excitation. At the same time there may be inhibition of a dilator center. Cessation of retinal stimulation and of excitation of the constrictor center, as in dark-adaptation, leads to decreased inhibition of the sympathetic tonus in the dilator center and contraction of the radial muscle. Release of the dilator center is accompanied by relaxation of the sphincter which may result from either decreased central excitation or from passive elastic stretch of sphincter fibers following cessation of excitation. In either case pupil dilatation in the normal eye accompanies the change in excitation level of the constrictor center during dark-adaptation. On this basis the similarity in the increase of pupil area and of cone sensitivity during dark-adaptation may be interpreted in terms of similar processes of recovery after constant excitation.

Summary

1. We have considered evidence that pupil size relative to light stimulation is under the control of fibers activated by cones in the central portion of the retina.

2. Apparatus and procedure used in photographing the pupil under infra-red radiation during dark-adaptation have been described.

3. The increase in pupil area is proportional to the increase in cone sensitivity during dark-adaptation following light-adaptation to 1200 ml.

(Manuscript received April 20, 1939)

References

1. Abelsdorff, G., Die Änderungen der Pupillenweite durch verschiedenfarbige Belichtung, *Z. Psychol. Physiol. Sinnesorg.*, 1900, **22**, 81–95.
2. Abelsdorff, G., and Feilchenfeld, H., Über die Abhängigkeit der Pupillarreaktion von Ort und Ausdehnung der gereizten Netzhautfläche, *Z. Psychol. Physiol. Sinnesorg.*, 1904, **34**, 111–131.
3. Bain, W. A., Irving, J. T., and McSwiney, B. A., The afferent fibers from the abdomen in the splanchnic nerve, *J. Physiol.*, 1935, **84**, 323–333.
4. Chance, B., Monochromatic vision, *Amer. J. Opthal.*, 1931, **14**, 417–425.
5. Crawford, B., Dependence of pupil size upon external light stimulus under static and variable conditions, *Proc. roy. Soc. Lond.*, 1937, **121B**, 376–395.
6. Engelking, E., Der Schellenwert der Pupillenreaktion und seine Beziehungen zum Problem der pupillomotorischen Aufnahmeorgane, *Z. Psychol. Physiol. Sinnesorg.*, 1919, **50**, 319–337.
7. Engelking, E., Vergleichende Untersuchungen über die Pupillenreaktion bei der angeborenen totalen Farbenblindheit, *Klin. Mbl. Augenheilk.*, 1922, **69**, 177–188.
8. Ferree, C. E., Rand, G., and Harris, E. T., Intensity of light and area of illuminated field as interacting factors in size of pupil, *J. exp. Psychol.*, 1933, **16**, 408–422.
9. Garten, S., Beiträge zur Kenntnis des zeitlichen Ablaufes der Pupillarreaktion nach Verdunkelung, *Pflüg. Arch. ges. Physiol.*, 1897, **68**, 68–94.
10. Geldard, F., The measurement of retinal fatigue to achromatic stimulation. *J. gen. Psychol.*, 1928, **1**, 578–590.
11. Geldard, F., The description of a case of total color blindness, *J. opt. Soc. Amer.*, 1933, **23**, 256–260.
12. Gullberg, J. E., Olmstead, J. M. D., and Wagman, I. H., Reciprocal innervation of the sphincter and dilator pupillæ, *Amer. J. Physiol.*, 1938, **122**, 160–166a.
13. Hecht, S., Nature of foveal dark adaptation, *J. gen. Physiol.*, 1921, **4**, 113–139.
14. Hecht, S., Haig, C., and Chase, A. M., The influence of light adaptation on subsequent dark adaptation of the eye, *J. gen. Physiol.*, 1937, **20**, 831–850.
15. Hertel, E., Experimenteller Beitrag zur Kenntnis der Pupillenverengerung auf Lichtreize, *Arch. Opthal.*, 1906, **65**, 106–134.

16. HESS, C., The extent of the pupillo-motor areas of the retina and pupillo-motor receiving organs, *Arch. Opthal.*, 1909, **38**, 131–146.
17. HESS, C., Farbenlehre, *Ergebn. Physiol.*, 1922, **20**, 1–107.
18. LAURENS, H., Studies on the relative physiological value of spectral lights. III. The pupillomotor effects of wave-lengths of equal energy content, *Amer. J. Physiol.*, 1923, **64**, 97–119.
19. LOHMANN, W., Über Helladaptation, *Z. Psychol. Physiol. Sinnesorg.*, 1907, **41**, 290–311.
20. LUCKIESH, M., AND MOSS, F. K., Area and brightness of stimulus related to the pupillary light reflex, *J. opt. Soc. Amer.*, 1934, **24**, 130–134.
21. MAGITÔT, A., *L'iris: étude physiologique sur la pupille et ses centres moteurs*, Paris: Doin, 1921.
22. MAGOUN, H. W., AND RANSON, S. W., The central path of the light reflex: a study of the effect of lesions, *Arch. Opthal.*, 1935, **13**, 791–811.
23. NUTTING, P. G., *Outlines of applied optics*, Philadelphia: Blakiston, 1912.
24. PURDY, D., Measurement of retinal illumination, *J. gen Psychol.*, 1932, **7**, 189–192.
25. REEVES, P., Rate of pupillary dilation and contraction, *Psychol. Rev.*, 1918, **25**, 330–340.
26. SACHS, M., Über den Einfluss farbiger Lichter auf die Weite der Pupille, *Pflüg. Arch. ges. Physiol.*, 1892, **52**, 79–86.
27. STEINACH, E., Untersuchungen zur vergleichenden Physiologie der Iris, *Pflüg. Arch. ges. Physiol.*, 1892, **52**, 495–525.
28. VERVOORT, H., Die Reaction der Pupille bei der Accommodation und der Konvergenz und bei der Beleuchtung verschieden grosser Flächen der Retina mit einer constanten Lichtmenge, *v. Graefes Arch. Opthal.*, 1900, **49**, 348–374.

GENERALIZATION AS A FUNCTION OF METHOD OF REINFORCEMENT

BY LLOYD G. HUMPHREYS

Department of Psychology, Institute of Human Relations, Yale University

Spence's (**17**) conditioned response theory of transposition, *i.e.*, use of generalization gradients to predict relational or absolute response, has recently been questioned by Razran (**14**). Perhaps the most telling argument advanced is that the positively accelerated [2] gradient assumed by Spence is not supported by the conditioning literature. If it can be shown, however, that evidence for Spence's gradient can be found in a simple conditioning experiment under suitable conditions, Razran's argument is considerably weakened.

One possibility of obtaining an initial positive acceleration in place of the usual negative acceleration is to test for generalization relatively near the point of reinforcement (**10**). Hovland's (**5**) tests for generalization in the pitch dimension, for example, were made at 25, 50, and 75 j.n.d.'s. The first step, then, is to extend Hovland's experiment to frequencies between 0 and 25 j.n.d.'s.

A second possibility is to use 50 percent reinforcement in place of the usual 100 percent reinforcement in the conditioning of the response; *i.e.*, reinforcement of only one-half of the acquisition trials in a random order. The writer (**8**) has shown that a greater resistance to extinction is established by this procedure in the conditioned eyelid response. It is also possible that generalization is affected. The 50 percent reinforcement situation is, at least, more analogous than 100

[1] The writer is indebted to the National Research Council for fellowship funds and to Professor Clark L. Hull under whose direction the work was done. Acknowledgment is also made to Professor W. R. Miles for the use of a sound-resistant room and to Professor D. G. Marquis for the galvanometer used in the investigation.

[2] Positive acceleration, as used in this connection, refers to increasing decrements of response corresponding to successive equal increments of dissimilarity between conditioned and generalized stimuli. For negative acceleration, on the other hand, the decrements become successively smaller rather than larger.

percent reinforcement to discrimination learning in as much as discrimination involves negative stimuli which are not reinforced.

Apparatus and Procedure

Hovland's apparatus and procedure for conditioning the psycho-galvanic response in human subjects (5) have been closely duplicated, with a few exceptions. The Burr-Lane vacuum-tube electrometer (2) was used in connection with a General Electric photo-electric, ink-recording galvanometer. It was thus possible to read responses in milli-volts; Hovland used millimeters. Hovland, by adjusting the sensitivity of his galvanometer, attempted to make the responses of all subjects of approximately equal amplitude. His standard deviations and standard errors of means are thus 'spuriously' low, but the standard errors of differences reported would, on the average, be unaffected, as artificial restriction of range is compensated for by reduced correlation coefficients.

In brief, the procedure was as follows. Two tones 5 and 15 j.n.d.'s removed from one of Hovland's reinforced tones were first determined. To these three was added Hovland's tone 25 j.n.d.'s removed. All four were then equated for loudness and for amplitude of response following equal reinforcement. In the experiment proper the four tones were first presented alone twice each, the shock (unpleasant, but not painful) was given twice alone, the 'reinforced' tone and the shock were paired 16 times for 100 percent reinforcement (8 pairings in 16 trials for 50 percent reinforcement), and then the four tones were again presented alone twice each to test for generalization. Trials occurred, on the average, once a minute.

With half of the subjects an additional variable was added in that the octave of the reinforced tone was used in place of the 25 j.n.d. tone. This served to test the importance of consonance vs. mere disparity in frequency as determiners of generalization.

In connection with the 'verbal-conditioning' of the psycho-galvanic response (3; 12), certain elements of the procedure deserve more detailed consideration. Initial responses to tones alone were recorded after the strength of the shock had been adjusted and after the electrodes were in place. Standard non-committal instructions were used, no mention being made of the pairing of tones and shock. It will be seen later that acquisition did occur.

Subjects were male college students, paid by the hour. None had previously served in a conditioning experiment.

Preliminary Standardization

There is considerable evidence (1) that the Weber fraction ($\Delta R/R$) is relatively constant between 1000 and 2000 d.v. It was thus decided to use the upper two of Hovland's frequencies, 1967 and 1000 d.v., for the 0 and 25 j.n.d. points, respectively, and to interpolate for 5 and 15 j.n.d.'s. These were computed to be 1718 and 1311 d.v., respectively. To these four frequencies was added 984 d.v., to be used as the octave of 1967.

In order to equate the chosen frequencies for loudness, the tone of 1000 d.v. was made the standard. Its intensity was set at 40 db. above threshold and equivalent to an attenuation box setting of 60 db. of attenuation. The results from 6 subjects for all frequencies expressed in decibels of attenuation relative to the standard tone and the settings which were used in subsequent experimentation are given below:

Frequency of tone	1967	1718	1311	1000	984
Mean attenuation	42.7	48.3	54.2	60.0	60.0
Experimental attenuation	42.0	48.0	54.0	60.0	60.0

The comparability of the present procedure with that of Hovland is indicated by the fact that similar results were obtained for 1967 d.v.

After equation for loudness the tones were presented, to 12 additional subjects, paired with shock and in rotated order as a further check on their comparability. The mean magnitudes of the conditioned responses obtained after a total of 16 reinforcements follow:

Frequency of tone	1967	1718	1311	1000 + 984
Mean response (in millivolts)	1.79	1.69	1.87	1.85

None of the differences are as large as their standard errors.

Generalization Following 100 Percent Reinforcement

Table 1 presents data, from 34 subjects, concerning re-

TABLE 1

Magnitude of Psycho-galvanic Responses Prior to, and Test-Trial Responses Following, 100 Percent Reinforcement of 1967 d.v. Data from 34 Subjects

Responses (in millivolts)	Frequency of Tone			
	1967	1718	1311	1000+984
Pre-reinforcement				
Median	1.25	.94	1.19	1.06
Mean	1.77	1.88	1.83	1.76
σ	1.90	1.97	1.93	2.26
Test-trials				
Median	3.38	2.71	2.81	2.50
Mean	3.52	2.91	2.84	2.81
σ	1.68	1.57	1.59	1.34

sponses to tones prior to, and test-trial responses following, 100 percent reinforcement of 1967 d.v. Computations are based upon means of responses for two trials. Responses to 1000 and 984 d.v. are, at present, combined.

Equality of response prior to reinforcement gives way to inequality of response following differential reinforcement. All of the differences between the reinforced response and the generalized responses are over three times their standard errors. A negatively accelerated gradient is suggested whether medians or means are considered.[3]

To state definitely the shape of a gradient when one's units are of doubtful equality or validity, however, is a dangerous procedure. It can be argued, for example, that millivolts of response are equal physically, but may not be psychologically; j.n.d.'s determined successively may not be valid units for generalization. There is evidence in this paper bearing on the latter point.

Validity of j.n.d.'s in Generalization

Consideration of responses to 1000 and 984 d.v. should furnish information on the validity of the j.n.d. units. Table 2 contains these data. Sixteen subjects under each condition

TABLE 2

MAGNITUDE OF PSYCHO-GALVANIC RESPONSES TO 1000 AND 984 d.v. BEFORE AND AFTER REINFORCEMENT OF 1967 d.v. 16 SUBJECTS IN EACH GROUP

	Pre-reinforcement Responses		Test-trial Responses		
	To Shock Mean	To Tone Mean	Mean	σ	t
1000 group.....	8.63	2.07	2.45	1.24	2.94
984 group.....	8.09	1.51	3.27	1.33	

were matched on the bases of presentation-order of test-trials and magnitude of the reinforced response (matching correla-

[3] Spence (18) plotted the medians for Hovland's (5) data and found an initial positive acceleration. Present agreement of means and medians from a larger number of subjects indicates that the previous discrepancy was probably the result of sampling error.

tions of 1.00 and .71, respectively). Responses to tones alone and to shock alone before pairing are included in the table to indicate the initial comparability of the two groups. For the mean difference between matched individuals in test-trial responses t is 2.94, indicating with a considerable degree of certainty that a true difference exists (P equals .01).

Successive j.n.d.'s of pitch are thus not the proper units to use for generalization, as consonance is involved. This factor probably influenced responses to other tones as well as to the octave, although it would be manifested in the latter to the greatest degree. The tone of 1311 d.v., for example, is to 1967 as 1 is to 1.5, *i.e.*, a consonant interval, while 1718 is distinctly dissonant.

Units of similarity should be substituted for j.n.d.'s. Unfortunately, however, generalization is probably the best criterion of similarity, although introspective judgments might be used as an independent criterion. On the other hand, it is quite possible that continua can be found where successive j.n.d.'s will be perfectly correlated with decreasing similarity. Intensity is probably such a continuum, but it possesses other complicating factors (**6**). Hues, on the contrary, would certainly exhibit differences between j.n.d.'s determined successively and degree of similarity.

Frame of Reference

A further complicating factor is that generalized responses may depend upon the setting in which the stimuli are perceived. In a sense 1967 is to 1000, in the present study, as 1967 was 153 in Hovland's study, *i.e.*, the highest and lowest of four tones presented. By making the reinforced response in each case equal to 100 percent, generalized responses can be put into comparable units. Responses to 1000 in the present study equal 72 percent of responses to 1967; responses to 153 in Hovland's study equal 70 percent, responses to 1000, 82 percent. This is suggestive evidence for the frame of reference notion, but a separate experiment would be needed to reach a definite conclusion. It is quite possible, for example, that degree of generalization is also dependent on the intensity

of the unconditioned stimulus (shock). There is no assurance that the shocks used in the two studies were equivalent.

In spite of the preceding limitations, the best guess one can make concerning the shape of the generalization gradient following 100 percent reinforcement is that of negative acceleration. The sharp drop to the 5 j.n.d. point and the small differences thereafter are not plausibly explained by errors in units; *e.g.*, such errors would have to be quite gross to convert a 'true' positive acceleration into the observed negative acceleration. Thus no support for Spence's hypothesis is found merely by using stimuli relatively close to the point of reinforcement (5, 15, and 25, j.n.d.'s).

Generalization Following 50 Percent Reinforcement

In Table 3 are included results, from 20 subjects, for responses to tones prior to, and test-trial responses following,

TABLE 3

Magnitude of Psycho-galvanic Responses Prior to, and Test-trial Responses Following, 50 Percent Reinforcement of 1967 d.v.
Data from 20 Subjects

Responses (in millivolts)	Frequency of Tone			
	1967	1718	1311	1000+984
Pre-reinforcement				
Median	1.31	.88	1.05	1.15
Mean	1.48	1.49	1.49	1.41
σ	.90	1.58	1.54	1.00
Test-trials				
Median	3.75	3.88	3.82	3.13
Mean	3.94	3.76	4.19	3.97
σ	2.09	1.97	2.50	2.94

50 percent reinforcement of 1967 d.v. The only difference in procedure between this group and the 100 percent reinforcement group is that shock followed the conditioned stimulus on only one-half of the 16 acquisition trials. Computations are based upon means of responses for two trials. Responses to 1000 and 984 d.v. are combined.

Equality of response prior to reinforcement is followed by approximate equality of response afterwards although 1967

d.v. was the only tone ever paired with shock. This is in striking contrast to the differential response obtained, following 100 percent reinforcement, from the first group of subjects.

Figure 1 shows the difference between the two conditions

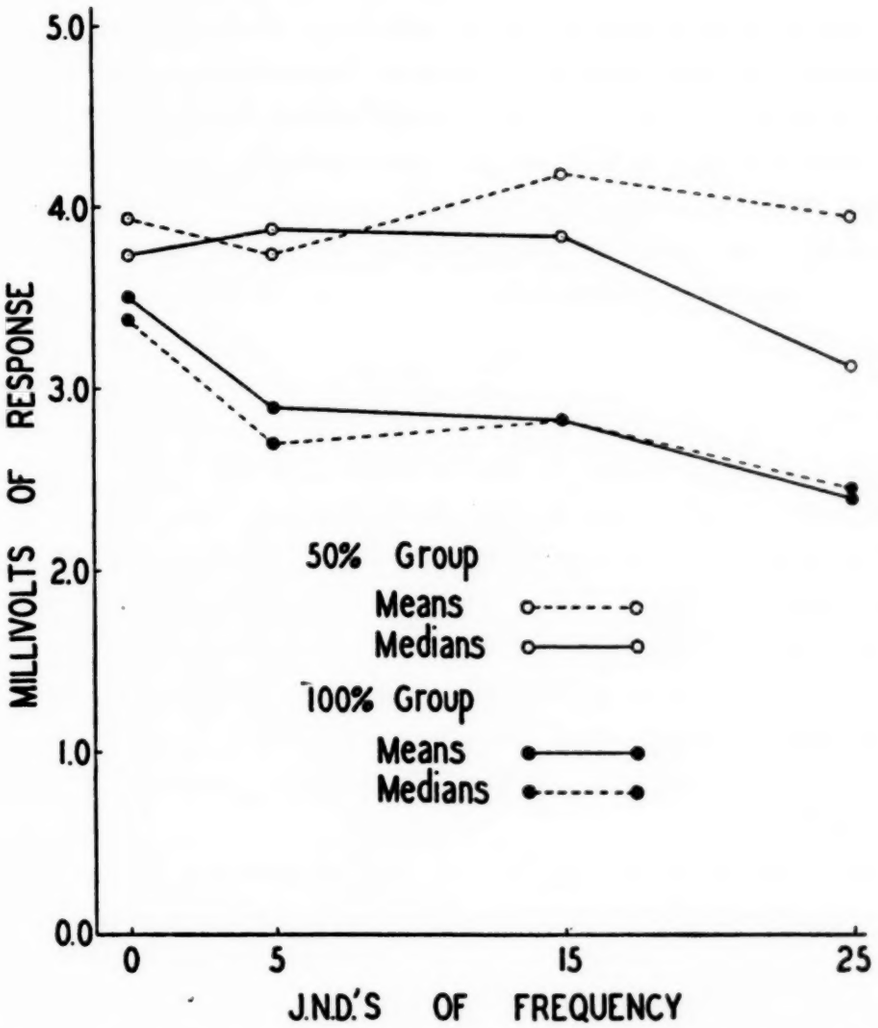

FIG. 1. Generalization as a function of method of reinforcement. Mean and median responses (in millivolts) are shown for the reinforced stimulus and for stimuli 5, 15, and 25 j.n.d.'s removed. Lines were drawn connecting the points for clarity Interpolation is undesirable. A negatively accelerated gradient following 100 percent reinforcement (34 subjects) and a positively accelerated gradient following 50 percent reinforcement (20 subjects) are indicated.

of reinforcement to better advantage. Both means and medians have been plotted. Responses to 984 d.v. in the 100 percent reinforcement group have been excluded from the 25 j.n.d. point because they differed markedly from those to 1000 d.v. (discussed in connection with Table 2).

The 100 percent reinforcement group exhibits the negative acceleration previously discussed, while the generalized responses of the 50 percent reinforcement group are essentially equivalent to each other and to the reinforced response. From this equivalence can be inferred an eventual phase of positive acceleration as it is unlikely that generalization is complete for the entire range of frequencies. Positive acceleration is, in fact, already manifested by the medians, but the drop at 25 j.n.d.'s is not statistically significant. This widespread, possibly positively accelerated generalization following 50 percent reinforcement may be relevant to Spence's theory of transposition.

Discussion

Before proceeding to the theoretical implications of the results, it is important to make sure that variables other than 50 percent reinforcement can be excluded from consideration.

The difference between 8 and 16 reinforcements could not be the responsible factor. Hovland (7) has shown that there is more generalization after 16 reinforcements than after 8 when reinforcement occurs on every acquisition trial. When 8 reinforced trials were interspersed with 8 non-reinforced trials in the present study, however, generalization was increased.

The difference in number of reinforcements under uniform 100 percent reinforcement conditions would have produced a difference in acquisition level (7) which might have been accentuated by the inclusion of non-reinforced trials with the smaller number of reinforcements. Contrary to this expectation, responses to the reinforced tone are higher in the 50 percent reinforcement group.[4] Thus the widespread generalization obtained in that group cannot be due to comparative failure of conditioning.

If a 'true' negatively accelerated gradient were forced against a physiological limit, a positively accelerated gradient would result. This condition would be revealed by decreased

[4] Greater resistance to extinction following 50 percent reinforcement is of considerable interest, but outside the scope of the present paper. It is discussed more fully elsewhere (9).

variability of response. Inspection of the standard deviations in Tables 1 and 3 indicates that this is not the case, as the 50 percent reinforcement group is actually more variable. The hypothetical physiological limit of psycho-galvanic responses is probably far above the present values.

More widespread generalization might be expected to follow from greater sensitivity to the conditioned and unconditioned stimuli used in the experiment. If the 50 percent reinforcement group gave larger responses to tone and to shock before pairing than the 100 percent reinforcement group, the difference in generalization might be explained. Contrary to this hypothesis, however, the former group was actually somewhat less sensitive. The means of responses to tones before pairing were 1.47 and 1.81, to shock, 7.30 and 8.61, respectively.

It can thus be concluded, with a reasonable degree of certainty, that 50 percent reinforcement is indeed responsible for the obtained results. The question then arises as to a possible relationship to discrimination and transposition.

Discrimination learning involves a negative stimulus which is not reinforced, and a positive stimulus which is always reinforced. Usually the positive and negative stimuli are presented an equal number of times in random order in establishing a conditioned discrimination. This is analogous to 50 percent reinforcement in simple conditioning. In the early stages in learning a trial-and-error discrimination, also, reinforcement occurs, on the average, on 50 percent of the trials. During over-learning, on the other hand, 100 percent reinforcement is received. In other words, 50 percent and 100 percent reinforcement in the simple conditioning situation represent the initial and final stages of discrimination learning.

On the basis of the preceding analysis the results from 50 percent and 100 percent reinforcement suggest an hypothesis concerning generalization gradients in discrimination. As the percentage frequency of correct choices increases, there will be a progressive change in the initial acceleration of the generalization gradient from positive to negative. This change in acceleration will be accompanied by progressive constriction in the amount of generalization.

The preceding hypothesis possesses several advantages over the simple positively accelerated gradient suggested by Spence (**17**). For one thing, a better basis is obtained for explaining very fine discrimination. The most efficient method to use in establishing such a discrimination is to start with stimuli easily discriminable and gradually reduce the difference between them (**13**). The result is that for a great many trials the subject has received 100 percent reinforcement. The gradient should thus be steeply negatively accelerated. The assumption of positive acceleration, on the other hand, is not plausible in this situation in that the subject would be required to react to very minute differences in excitatory strength.

Recent results (**11**; **15**) indicate that the range of relational responses decreases and the number of absolute choices increases late in discrimination learning. Razran (**14**) states that this is contrary to conditioning facts, constriction of generalization being a phenomenon which occurs in earlier stages. On the basis of the present hypothesis, however, one would expect the range of relational responses to be greatly reduced *only* after some degree of overlearning had occurred.

A question as to the generality of results obtained with the psycho-galvanic response in human subjects may be raised because of the possibility of so-called verbal conditioning (**3**; **12**). The results should, of course, be checked with lower organisms and with other responses, but it is not felt that the possibility of verbal conditioning places the present research in a class by itself. Conditioning without paired stimulation has been demonstrated in animals other than man and with responses other than the psychogalvanic (**4**; **16**), so the phenomenon is not necessarily verbal. In addition, numerous findings indicate that preparatory set or expectancy is an important factor in most 'true' conditioning.

Differences in generalization following 50 percent and 100 percent reinforcement are, in fact, explicable in terms of expectancy. Introspective comments of the subjects indicated that little difficulty was experienced in forming an expectancy of continuous non-reinforcement in the test-trials following 100 percent reinforcement and that frequency differences were

plainly apparent. Greater difficulty in forming the same expectancy was experienced following 50 percent reinforcement, and failure to report frequency differences was more common. It is not unreasonable to assume that subjects in the latter situation were so much more concerned about the possible occurrence of the unconditioned stimulus (shock) that frequency differences went unnoticed.

Summary and Conclusions

Generalization of conditioned psycho-galvanic responses to tones 5, 15, and 25 j.n.d.'s distant from the reinforced tone has been investigated under two conditions of reinforcement. One group of 34 human subjects was given 100 percent reinforcement for 16 trials, every trial being reinforced, followed by 8 non-reinforced test-trials. An additional group of 20 subjects was given 50 percent reinforcement for 16 trials, only one-half of the trials being reinforced, followed by 8 non-reinforced test-trials. With half the subjects the octave of the reinforced tone was presented in place of the 25 j.n.d. tone to determine the importance in generalization of consonance vs. mere disparity in frequency. From the results the following conclusions can be drawn.

1. Responses to the reinforced tone are significantly higher than generalized responses following 100 percent reinforcement. A negatively accelerated gradient is suggested by both means and medians.

2. Complete generalization, measured either by means or by medians, is obtained following 50 percent reinforcement. This suggests a phase of positive acceleration.

3. Responses to the octave of the reinforced tone are significantly higher than responses to the 25 j.n.d. tone, although the former is approximately 26 successive j.n.d.'s removed. Thus gradients of generalization should not be based upon j.n.d. units unless there is evidence that these units are perfectly correlated with decreasing similarity.

Conclusions 1 and 2 form the basis for the following hypothesis: as the percentage frequency of correct choices in discrimination learning increases, there will be a progressive change in the initial acceleration of the generalization gra-

dients from positive to negative, accompanied by progressive constriction in the amount of generalization. Some of the implications of this hypothesis for Spence's theory of transposition are pointed out.

(Manuscript received April 19, 1939)

References

1. Banister, H., Audition. I. Auditory phenomena and their stimulus correlations, in Murchison, C., *Handbook of general experimental psychology.* Worcester: Clark Univ. Press, 1934, 880–923.
2. Burr, H. S., and Lane, C. T., Electrical characteristics of living systems, *Yale j. biol. med.*, 1935, **8**, 31–35.
3. Cook, S. W., and Harris, R. E., The verbal conditioning of the galvanic skin reflex, *J. exper. Psychol.*, 1937, **21**, 202–210.
4. Grether, W. F., Pseudo-conditioning without paired stimulation encountered in attempted backward conditioning, *J. comp. Psychol.*, 1938, **25**, 91–96.
5. Hovland, C. I., The generalization of conditioned responses. I. The sensory generalization of conditioned responses with varying frequencies of tone, *J. gen. Psychol.*, 1937, **17**, 125–148.
6. ——, The generalization of conditioned responses. II. The sensory generalization of conditioned responses with varying intensities of tone, *J. genet. Psychol.*, 1937, **51**, 279–291.
7. ——, The generalization of conditioned responses: IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses, *J. exper. Psychol.*, 1937, **21**, 261–276.
8. Humphreys, L. G., The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions, *J. exper. Psychol.* (in press).
9. ——, Extinction of conditioned psycho-galvanic responses following two conditions of reinforcement (to appear).
10. Hull, C. L., The problem of stimulus equivalence in behavior theory, *Psychol. Rev.*, 1939, **46**, 9–30.
11. Jackson, T. A., Stonex, E., Lane, E., and Dominguez, K., Studies in the transposition of learning by children. I. Relative vs. absolute response as a function of the amount of training, *J. exper. Psychol.*, 1938, **23**, 578–600.
12. Mowrer, O. H., Preparatory set (expectancy)—a determinant in motivation and learning, *Psychol. Rev.*, 1938, **45**, 62–91.
13. Pavlov, I. P., *Conditioned reflexes*, London: Oxford Univ. Press, 1927.
14. Razran, G. H. S., Transposition of relational responses and generalization of conditioned responses, *Psychol. Rev.*, 1938, **45**, 532–538.
15. ——, Studies in configural conditioning. V. Generalization and transposition, *J. educ. Psychol.*, 1939 (in press).
16. Sears, R. R., Effect of optic lobe ablation on the visuo-motor behavior of goldfish, *J. comp. Psychol.*, 1934, **17**, 233–265.
17. Spence, K. W., The differential response in animals to stimuli varying within a single dimension, *Psychol. Rev.*, 1937, **44**, 430–444.
18. ——, A reply to Dr. Razran on the transposition of responses in discrimination experiments, *Psychol. Rev.*, 1939, **46**, 88–91.

STUDIES IN THERMAL SENSITIVITY: 12 PART-WHOLE RELATIONS IN SERIATIM COLD-MAPPING

BY WILLIAM LEROY JENKINS

Lehigh University

In a preliminary study of seriatim cold-mapping with stimulators of different sizes,[1] it has been shown that simple summation is not the rule; *i.e.*, that doubling the stimulus area does not yield twice as large a seriatim score. In the present investigation, to determine more accurately the relation between stimulus area and seriatim scores, the same cutaneous regions have been mapped alternately with two stimulators, one being one-half or one-third the area of the other.

Procedure

The method of seriatim mapping was similar to that employed in the preceding study.[2] Untrained student subjects worked in groups of three, one acting as experimenter, the other two alternating as subject and recorder. A checkerboard pattern of suitable squares or rectangles, stamped on the volar surface of the left forearm, was mapped 12 times—6 rounds with each size.[3] The report categories were: 0 for neutral, 1 for weak, 2 for medium and 3 for strong cold. Inlet temperature was maintained at 17° C. in all experiments.

Figure 1 shows some sample data from one row of squares (10 out of 50) in the series employing 3 × 3 and 3 × 6 stimulators alternately. The upper blocks give the successive reports from 3 × 6 stimulation; the lower those from 3 × 3 stimulation. Subjects do not always respond with exactly the same number to repeated stimulation of the same square. Occasionally, a completely erratic report will appear—possibly the result of misplacement of the stimulator by the experimenter or of inattention on the part of the subject. The simple sum of the numbers is always taken as the *seriatim*

[1] Jenkins, W. L., Studies in thermal sensitivity. 11. Effects of stimulator size in seriatim cold-mapping, *J. exper. Psychol.*, 1939, **25**, 302–306.

[2] *Op. cit.*

[3] Consider the 3 × 3—3 × 6 series as an example: The sequence was: 3 × 3 on first subject; 3 × 3 on second subject; 3 × 6 on first; 3 × 6 on second, etc. until 6 rounds with each size had been mapped on each subject. A minimum time of 3½ min. per round was rigidly maintained, to equalize adaptation effects. The stimulation followed approximately a random order to avoid any possibility that the subject could memorize a sequence of reports.

3 × 6	2	0	3	3	1
	3	0	3	3	1
	2	0	3	3	1
	3	0	3	3	1
	3	1	3	2	1
	2	0	3	2	1

3 × 3	2	2	1	0	2	3	2	3	1	2
	1	1	0	1	3	3	3	3	2	1
	1	2	1	1	2	2	3	2	2	0
	2	3	2	1	2	3	3	2	1	0
	2	2	0	0	2	2	2	3	1	0
	3	3	2	1	3	3	3	2	1	1

FIG. 1. Excerpt from typical data sheets.

score, however, without any attempt at arbitrary correction. This summing of ordinal numbers as if they were cardinal is theoretically reprehensible, but appears to be empirically justified in this case by the consistent relations among the scores thus computed. In practice, 0, 1, 2 and 3 must approach a linear series for such relations to hold.

RESULTS

Table 1 gives both the reliability coefficients from repetition with the same size and the coefficients of correlation between whole scores and mean part scores when different sizes

TABLE 1

From Repetition with the Same Size

Stim. Size	.00 .59	.60 .64	.65 .69	.70 .74	.75 .79	.80 .84	.85 .89	.90 .94	.95 1.00	Total Cases	
3 × 3	0	2	3	5	2	7	8	3	1	31	
3 × 1½	1	0	1	3	5	3	6	1	0	20	52%
2 × 2	11	4	6	8	14	13	20	6	0	82	over
1½ × 1½	2	0	3	2	2	5	6	1	0	21	.80
1 mm	0	2	3	2	6	2	4	5	4	28	
Between Total Scores and Average Part Scores											
3 × 6 3 × 3	1	2	2	6	5	8	1	7	0	32	
3 × 3 3 × 1½	2	2	1	5	5	6	4	2	1	28	50%
2 × 4 2 × 2	3	1	2	2	3	5	6	4	0	26	over
3 × 1½ 1½ × 1½	1	0	3	3	2	1	6	2	0	18	.80
2 × 2 2 × 1	4	1	0	1	4	3	4	3	0	20	
2 × 1 1 mm	7	1	4	2	5	6	1	2	0	28	
3 × 4½ 3 × 1½	4	0	0	1	1	1	4	1	0	12	
2 × 6 2 × 2	2	1	4	3	3	8	8	3	0	31	

Figures in the table indicate the number of cases within the given step interval.

are used alternately. In both instances, these are .80 or higher in approximately 50 percent of the cases, indicating some extremely orderly relation between whole and part scores. To avoid any possibility that the results might be distorted by random reporting, only the data from those subjects with correlations of .80 and over have been used in the ensuing analyses. (As a matter of fact, almost identical curves are obtained by using all the data indiscriminately.)

The Gross Relation to Stimulator Size.—By combining the data for all reliable subjects, we obtain Fig. 2, in which average

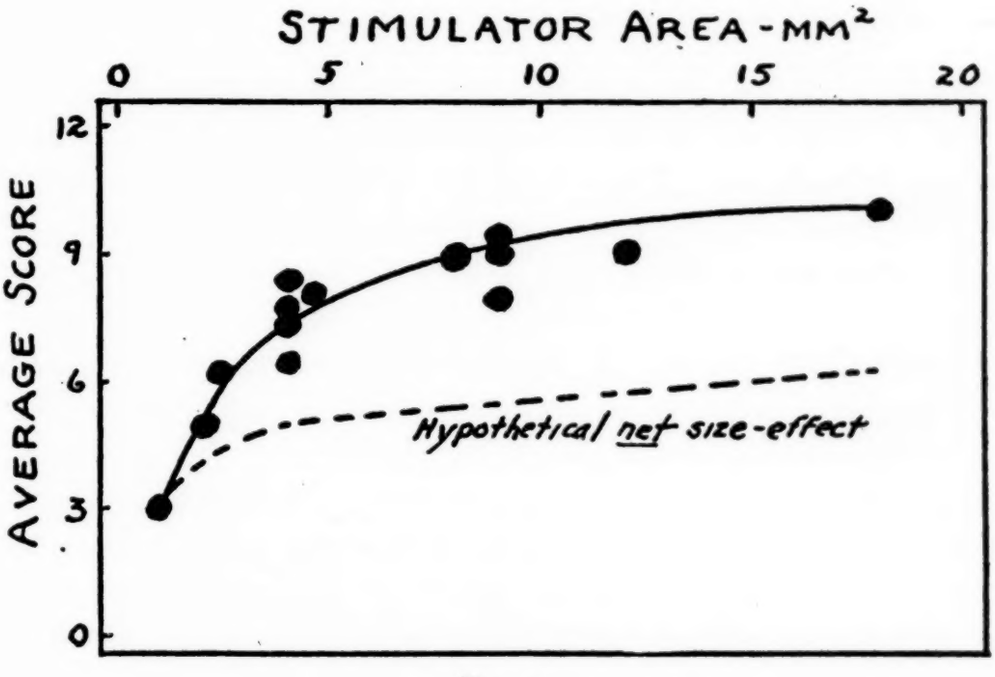

FIG. 2

scores are plotted against stimulator area, the circles indicating average scores for different groups of 10 or more subjects, and the solid line approximately a best-fitting curve. This is not intended as a quantitative representation, but sets the stage for the next question:

To what extent is this curve a function of stimulator area as such? The answer must be approached indirectly.

The Hypothetical Net Size-Effect.—Table 2 shows whole and part average scores from alternate mapping—selecting only those cases where a whole score can be matched with two *identical* part scores. The percentage increase of whole over

TABLE 2

Part Size	Whole Size	Mean Part Score	Mean Whole Score	Increase: Whole over Part
Square-to-Rectangle				
3 × 3	3 × 6	9.1	9.8	8%
2 × 2	2 × 4	8.0	8.2	1%
1½ × 1½	3 × 1½	3.7	4.1	11%
Rectangle-to-Square				
3 × 1½	3 × 3	5.0	6.0	20%
2 × 1	2 × 2	2.8	3.4	21%

part score in such instances represents the effect of stimulator area as such, divorced from any possible influence of part score differences. The increases range from 1 percent to 21 percent.[4] By applying these percentages successively, we can construct a hypothetical curve which should represent the *net* influence of stimulator area. This is shown as the dotted line in Fig. 2. A liberal allowance has been made for the increase from 1 to 2 mm², because the average score figures at the smallest size are so overweighted with double-zeros as to make direct comparison impossible. The hypothetical net size-effect curve [5] clearly falls far below the solid line in Fig. 2 which represents the gross effect of stimulator size. The implication is plain: Some factor other than stimulator size as such must operate. This factor must be related to part score *differences*.

The Basic Curve.—As a preliminary step toward identifying this function, we must abandon averages and investigate the relation between individual whole and part scores. The upper left-hand graph of Fig. 3 results from plotting individual

[4] The square-to-rectangle increase is consistently less than that for rectangle-to-square, which implies that rectangular stimulation is 'weaker.' The effect of stimulator form and stimulator orientation are topics for further investigation.

[5] Theoretically, by mapping an area successively with larger and larger stimulators and selecting the ever-diminishing number of cases showing identical part scores at all size levels (indicating uniform sensitivity), it would be possible to develop the net size-effect curve empirically. Since we are interested in the curve only provisionally, this time-consuming task has not been undertaken.

whole scores against individual part scores, selecting those cases where *identical* parts can be matched with whole scores. For this purpose, all such cases from the six whole-half comparisons have been combined, neglecting the variations in percentage of increase shown in Table 2. The circles indicate the medians of whole scores, with open circles where these are

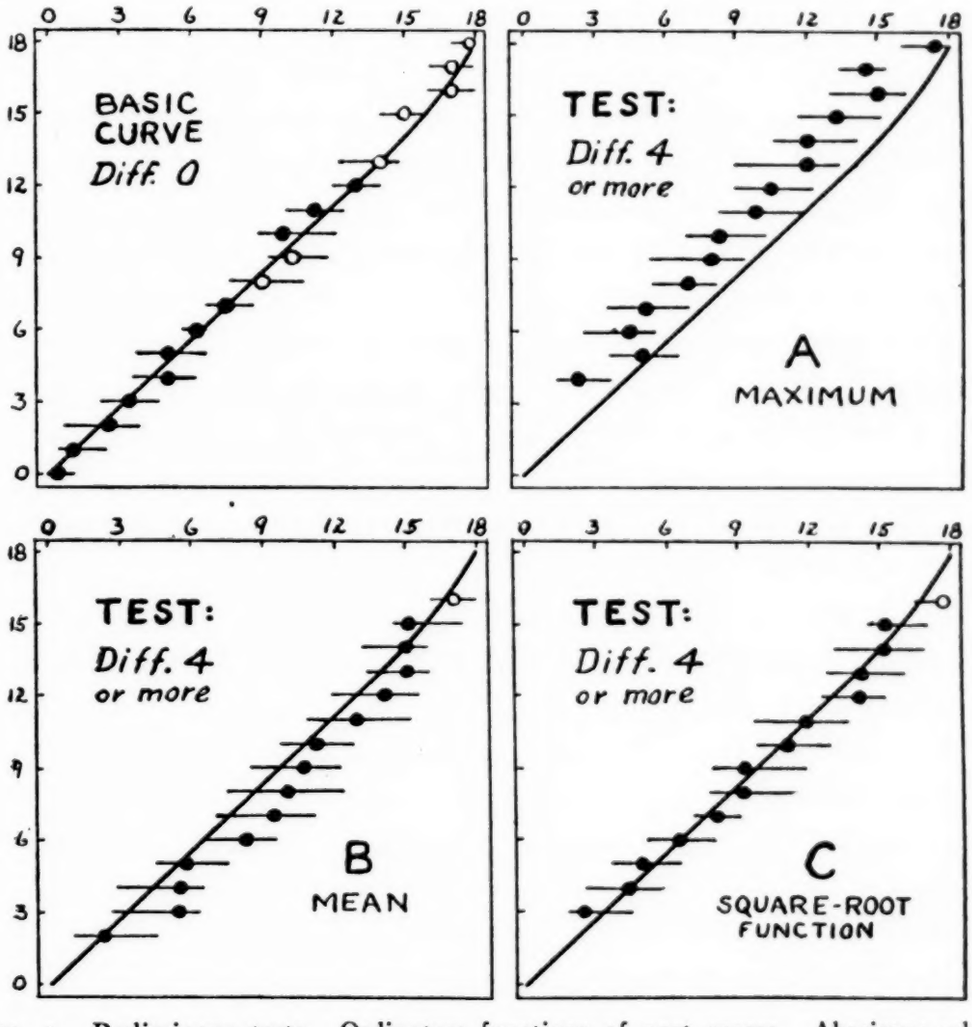

FIG. 3. Preliminary tests. Ordinates: functions of part scores. Abscissæ: whole scores. Circles: medians of whole scores. Lines: interquartile ranges.

based on less than 10 cases. Horizontal lines indicate the interquartile ranges.[6]

[6] The small size of the quartile deviation—less than 1½ units generally—provides empirical justification for the method of determining scores by adding the category numbers reported. Although there are wide individual differences in score levels—although each subject sets his own standards for what he will call 0, 1, 2 and 3—there is evidently a common basis in the *relationship* between 0 and 1, 1 and 2, and 2 and 3 as they are employed by most subjects.

The heavy line in this graph represents approximately a best-fitting curve. The slope of the linear portion indicates an average 8 percent increase in whole over part scores, which agrees substantially with the smaller average score increases in Table 2 and provides further justification for our hypothetical net size-effect curve of Fig. 2.

Preliminary Tests.—The basic curve has been plotted only from cases where the part scores are identical. The next step is to discover a function of the part scores which provides a good fit when the components are *non-identical.* Since the true relation will be revealed more sharply by cases with large differences, only those where part scores differ by 4 or more units are used in the preliminary tests shown in the remaining graphs of Fig. 3.

The *maximum* (*A*) of part scores is clearly not the correct function. Whole scores are lower than maximum part scores. The *mean* (*B*) is somewhat better, but whole scores are higher than mean part scores.[7] The *square-root function* (*C*) provides a fit to the basic curve which is approximately as good as that of the identical score medians from which it was derived. This function is the square root of the mean of the squares of the part scores (mathematically similar to the formula for standard deviation).

Critical Checks.—Two checks for this function immediately suggest themselves:

1. Is the fit to the basic curve good with all magnitudes of part score differences? Figure 4 provides the answer. With differences of 1–2–3 and 4–5–6, the coincidence is excellent. With differences of 7 or more, six of the medians touch the curve, while four are about one score point higher. It is dangerous to explain away such discrepancies as 'experimental errors,' because further investigation may show that some modification of the function is necessary. Nevertheless, the following opportunity for artifacts must be noted: Where large differences in part scores exist, it means that the level of sensitivity is changing sharply. Consequently, even slight

[7] In any event, the mean would be a poor choice since it gives no chance for an influence of part score differences.

misplacements of the stimulator will lead to erratic reports. In a region where there is a *general* predominance of high sensitivity, it is conceivable that misplacements into regions of higher sensitivity might occur more frequently than into regions of lower sensitivity. This would automatically make whole scores a trifle higher than the basic curve calls for.

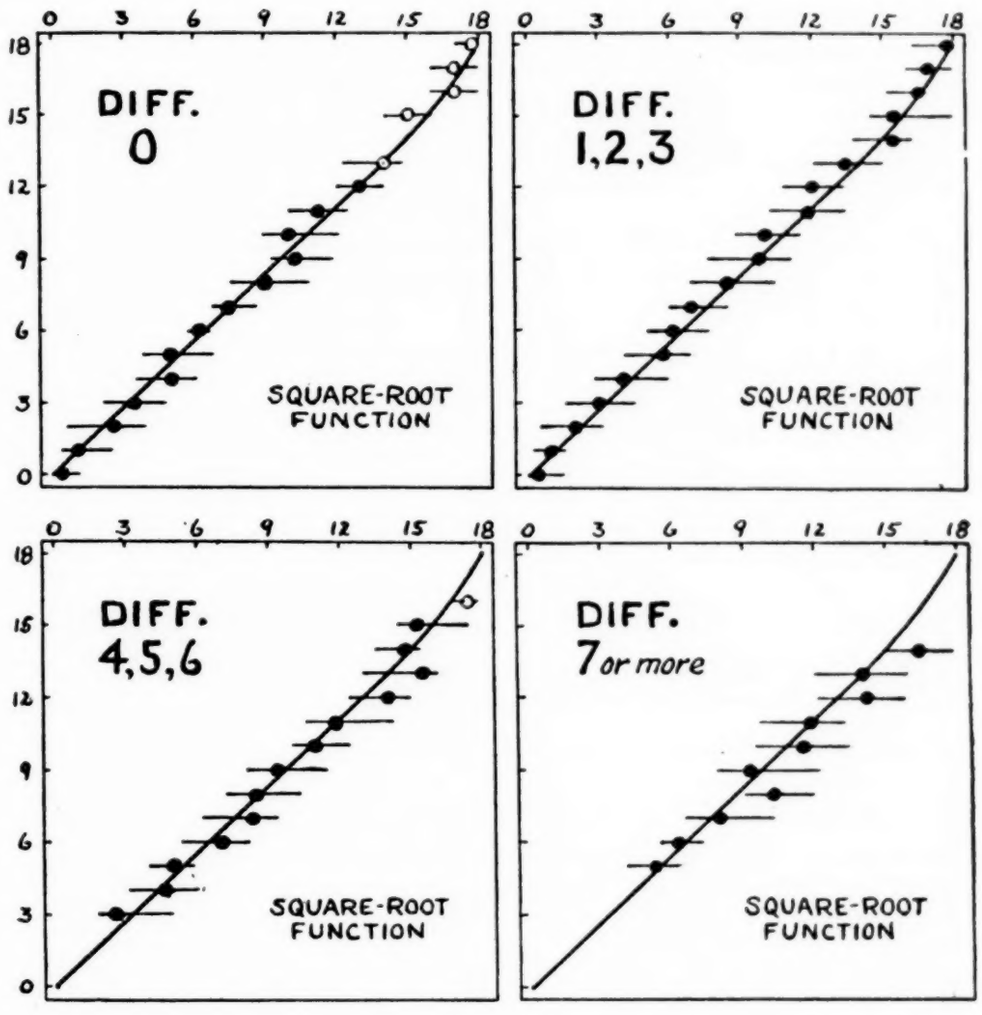

FIG. 4. Square-root function applied to various part score differences. Ordinates: square-root function of part scores. Abscissæ: whole scores. Circles: medians of whole scores. Lines: interquartile ranges.

2. Is the fit to the basic curve good with all size comparisons? Figure 5 shows that it is. Even the slight departures to the right or left of the basic curve agree with the varying percentage increases shown in Table 2, which was based on identical part scores. Thus, the medians for $3 \times 4\frac{1}{2}$

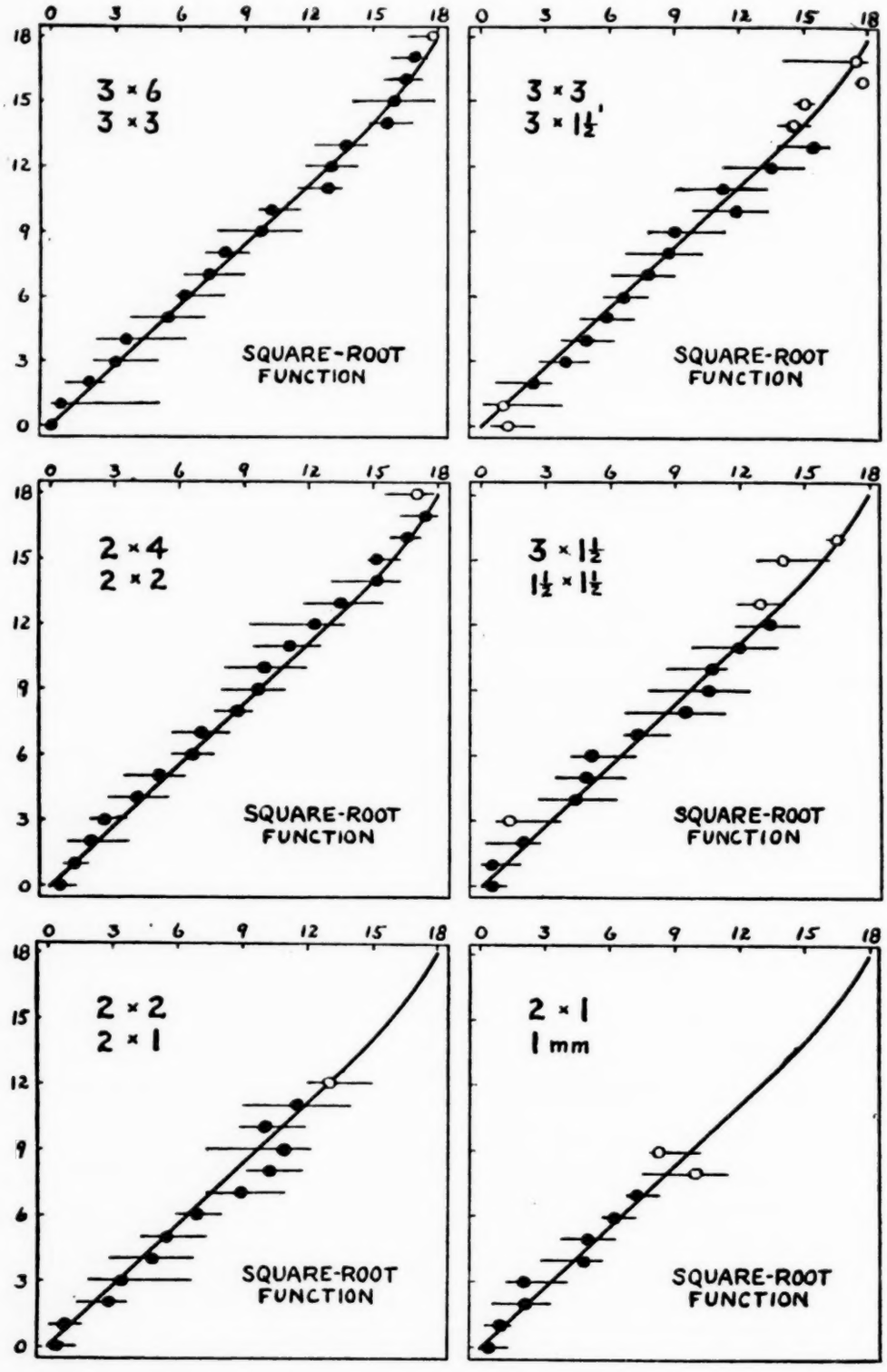

FIG. 5. Square-root function applied to whole-half comparisons. Ordinates: square-root function of part scores. Abscissæ: whole scores. Circles: medians of whole scores. Lines: interquartile ranges.

$-3 \times 1\frac{1}{2}$ and $2 \times 2 - 2 \times 1$ (20% and 21% in Table 2) veer to the right, while the medians for $2 \times 4 - 2 \times 2$ (1%) are slightly to the left, and those of $3 \times 6 - 3 \times 3$ (11%) are almost perfectly centered. Confirmation is furnished by Fig. 6 where two whole-third comparisons are presented. We have no basic curve for whole-third relations, because cases with three identical components are rare. However, the fit to the whole-half basic curve is very good, especially for the $2 \times 6 - 2 \times 2$ series.

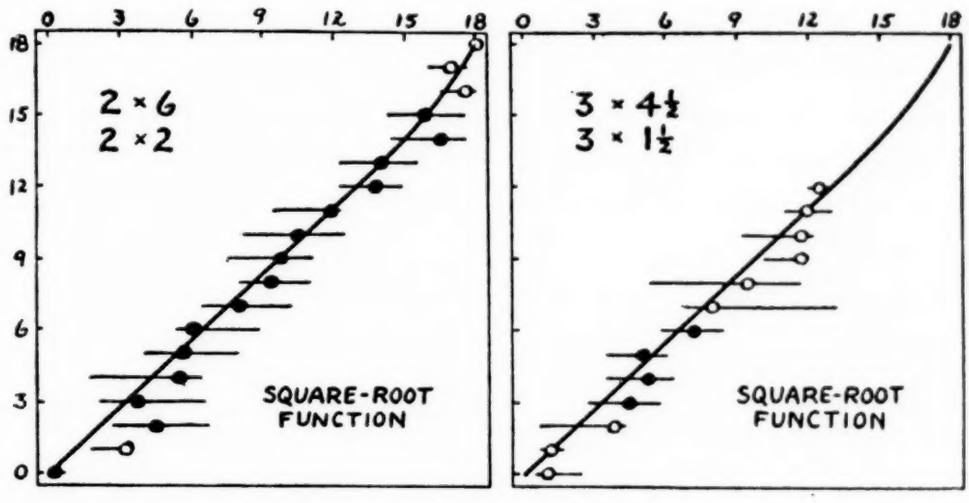

FIG. 6. Square-root function applied to whole-third comparisons. Ordinates: square-root function of part scores. Abscissæ: whole scores. Circles: medians of whole scores. Lines: interquartile ranges.

From the results of these critical checks, it appears that the square-root function holds within the range of stimulator sizes investigated. If not an exact mathematical statement of the relation between whole and part scores (when size-effect as such is excluded), it is certainly a satisfactory first approximation.

DISCUSSION

'Cold Spots' Inadequate.—Since the pioneer studies of Blix, Goldscheider and Donaldson, it has been generally held that cutaneous sensitivity is punctiform. In a strictly descriptive sense, this merely means that sensory 'spots' can be mapped on the skin with small stimulators. However, if such 'spots' are to have any intrinsic significance—if they are

to be more than incidental artifacts of a particular method of mapping—they must represent ultimate units corresponding to receptors. It has been generally assumed that the 'spots' *do* correspond to receptors. Thus, we have the von Frey hypothesis that 'cold spots' are Krause end-bulbs—a view which Strughold, Bazett and others have attempted to support with histological evidence. Even those who dispute the von Frey theory have generally sought for some kind of receptors corresponding to mapped 'cold spots' as units.

Seriatim mapping should be an ideal procedure for localizing 'cold spots,' because it provides statistical evidence of the reliability of reporting, which is not possible with single mapping. Isolated positive squares, surrounded by consistent zero squares, should be a feature of seriatim maps made with a 1 mm diam. stimulator—the traditional size for mapping 'cold spots.'

How does this theory fit the experimental facts? Figure 7 shows seriatim 1 mm maps from 10 highly consistent subjects—reliability coefficients ranging from .86 to .98. The left-hand maps give the actual seriatim scores. In spite of wide individual differences in general level, these maps show one common feature: *There are no discrete 'cold spots.'* There are consistent positive squares and consistent zero squares, but no isolated positive squares surrounded by zero squares, except in a few dubious instances. As a general rule, *the positive squares are grouped in clusters.*

It is idle to suggest that each square in a cluster represents a separate 'cold spot.' There is nothing sacred about these particular squares—the result of an arbitrary pattern. Stamping a slightly different pattern would give a new set of squares and require the assumption of a new set of 'cold spots' to fit. The number of 'cold spots' would then be indeterminate.

The Resolving Limits of Mapping.—Logically, the next step is to try smaller and smaller stimulators until resolution of the clusters into their component units is attained. Practically, this involves a number of difficulties. With a 1 mm diam. stimulator, we have reached about the limit of accurate

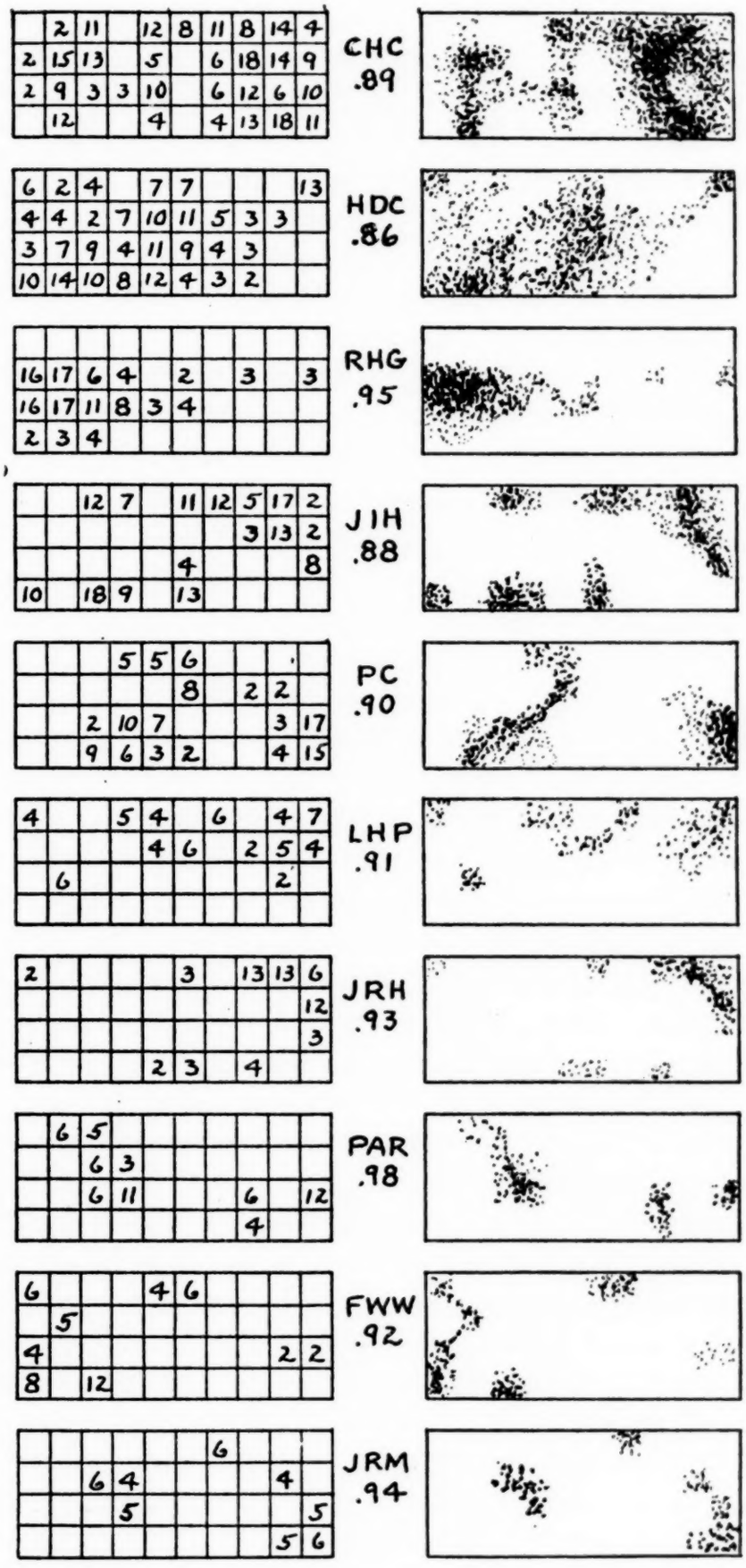

FIG. 7. Seriatim maps with 1 mm diam. stimulator. Each square in maps at left represents 1 mm². Maps at right illustrate the concentration theory.

manual placement. This might be overcome by the use of mechanical stimulation, if we could find some way of immobilizing the skin for a two-hour mapping period—no small task in itself. Again, the heat capacity of very small stimulators might set a limit on usable size. Even more important, the *spread* of stimulation (roughly proportional to the perimeter) might defeat our efforts to obtain a stimulator *functionally smaller* than a certain size.

Even if we could overcome these technical difficulties, however, there is nothing in the available evidence to make us expect that resolution can ever be attained. The maps in Fig. 7 are arranged in descending order of the number of positive squares. The upper two resemble those usually obtained with larger stimulator sizes, and the descending series roughly recapitulates the change in the appearance of the maps which occurs when we go from 3 × 3 to 2 × 2 to 2 × 1 to 1 mm diam. As we go down the series of maps, it is evident that the percentage of positive squares is becoming less and less—yet the remaining positive squares are still grouped into clusters. The lower maps are simply *disintegrating without being resolved into ultimate units.* Extrapolation of the process offers little hope that the end-point will be anything more than artifacts. Conceivably, we might reduce the stimulator size until only a single tiny point remains responsive out of each cluster. This, however, would not be a resolution of the cluster into component units, but merely a delimitation of the 'peak' of the cluster—no more indicative of the topography of the region than the last church spire which remains visible above flood waters.

Thermal-Conduction Inadequate.—Such 'peaks' could possess intrinsic significance only if each corresponded to a single receptor which is responsible for the variations of sensitivity found by mapping with somewhat larger stimulators. Thus, a *cluster* might represent a single receptor (or several of them)—the surrounding gradations being determined by thermal-conduction. Hardy and Oppel[8] have

[8] Hardy, J. D. and Oppel, T. W., Studies in temperature sensation. III. The sensitivity of the body to heat and the spatial summation of the end-organ responses, *J. clin. Investig.*, 1937, **16**, 533–540.

suggested such a mechanism for warmth, thermal conduction to the receptor occurring through a network of blood vessels, making the receptor effective over much more than its actual area. This could account for two different part scores with a single end-organ as the intermediary. The critical test is the relation of whole to part scores. When a larger stimulator is applied to cover both parts simultaneously, in terms of the thermal-conduction hypothesis, the resulting intensity must be at least that of the maximum of the parts. A single receptor which is responsible for two part scores cannot conceivably be stimulated less strongly by a stimulator which covers both parts.

Yet this is exactly what does happen experimentally! The whole score, neglecting the minor size-effect, is the *square-root function* of the differing part scores, which is mathematically always *less* than the maximum. The establishment of the square-root function effectively shuts out thermal-conduction as an explanation of the variations in sensitivity found in seriatim maps. Hence, neither the traditional 'cold spots' nor thermal-conduction to isolated receptors can describe the distribution of sensitivity to cold.

The Concentration Theory.—How can we account for a distribution which is non-uniform, yet neither mappably punctiform, nor dependent upon variations in thermal-conduction? Let us assume that there are a great many minute receptors in a given skin area. These may vary individually in threshold and other characteristics—but fundamentally their *concentration* determines the sensitivity of the area. Where the receptors are densely packed, there is high sensitivity; where they are sparse, there is little or none. The situation is crudely represented by the shaded maps on the right-hand side of Fig. 7, deeper shading indicating a heavier concentration of receptors.

In terms of this theory, the intensity of cold experienced from a stated temperature is determined by two factors:

1. The *sensitivity* of the stimulated area—proportional to the square-root function of the concentration, or mathematically, the square root of the total number of receptors divided

by the area. With stimulators approximately 2 mm^2 and larger, this, and not size, is the primary determinant.

2. The *size* of the stimulated area—proportional mathematically to the total number of receptors stimulated. This is relatively unimportant with larger stimulators, but may exercise considerable influence below about 2 mm^2.

The theory, as above stated, is somewhat primitive. We may hope ultimately to reduce it to more basic neurological concepts; for example to the frequency of nerve impulses. Thus, the sensitivity factor might be translated into the *square root of the total frequency per unit area*, where 'total frequency' means the number of nerve impulses per second reaching the central nervous system from the area. The size factor would then become some fraction of the total frequency itself. We should like also to be able to identify the hypothetical minute receptors with specific anatomical entities; for example, with some of the 'free nerve endings,' usually assigned to pain by a process of exclusion. We might even hope eventually to discover some sort of differentiation which would constitute a distinguishing feature for those subserving cold sensitivity. All of these are goals for further research.

Broader Implications.—These findings with cold immediately raise the question: Is warm sensitivity punctiform in the traditional sense? Do 'warm spots' designate unit receptors any better than 'cold spots'? Seriatim maps of warm sensitivity—made with 2 and 3 mm stimulators—are similar in appearance to those for cold.[9] It seems equally improbable that the distribution can be adequately explained on the basis of 'warm spots'—especially as the number usually given is only one or two per cm^2. Thermal conduction from a single receptor cannot be used to explain the variations any more than with cold, unless it turns out that whole scores are equal to or greater than the maximum of part scores. Part-whole relations in seriatim warm-mapping will provide the critical evidence.

[9] Jenkins, W. L., Studies in thermal sensitivity. 10. The reliability of seriatim warm-mapping with untrained subjects, *J. exper. Psychol.*, 1939, **24**, 439–449.

Finally, some of the results suggest a possible extension beyond the field of temperature sensitivity. Four broad questions might be proposed:

1. Is the concentration theory a fundamental rule for all forms of cutaneous sensitivity?

2. Are all cutaneous receptors to be found among the 'free nerve endings,' with terminal differentiations for the various modalities?

3. Can the square-root function be extended into a general neurological law? That is, is the intensity of experience fundamentally proportional to the *square root of the total frequency of nerve impulses per unit area* in all sensory modalities—a modification of Adrian's frequency hypothesis? (In addition, there would be the minor size-effect—some fractional function of total frequency itself, a modification of Lucas' 'number of neurons active.')

4. Would the square-root function, if found translatable into nerve impulses, express a psychophysical relation of the kind Fechner was seeking? In place of measurements of external physical energy applied, we could employ the more direct stimulus units: frequency of nerve impulses. For the dubious direct measurement of sensation units, we could substitute seriatim scores of known reliability. The square-root function would then correspond to the Fechnerian concept of a logarithmic relation between stimulus and sensation.

These broad problems run far beyond the limits of the present data and are presented solely as a spur to further research.

Summary

Seriatim cold-mapping of the same cutaneous region alternately with two stimulators, the smaller being one-half or one-third the area of the larger, shows that scores indicating intensity of experience are dually determined: (1) There is a minor effect of size as such, relatively unimportant except with very small stimulators. (2) The major influence is the *square-root function*—whole scores being equal to the square root of the mean of the squares of differing part scores. Criti-

cal checks show that this square-root function holds with all magnitudes of part score differences and with all size comparisons tested.

Highly reliable seriatim maps with a 1 mm diam. stimulator reveal no isolated 'cold spots'; *i.e.*, positive squares surrounded by zero squares. The positive squares are always grouped in clusters, even in maps where very few of them remain. It is maintained that mapping with smaller stimulators cannot resolve these clusters into ultimate units, because the maps are simply disintegrating with no sign of resolution. The gradations within a cluster cannot be explained on the basis of varying thermal-conduction to a single receptor, because this would require whole scores equal to or greater than the maximum of part scores, whereas the relation follows the square-root function (necessarily less than the maximum). Neither the traditional 'cold spot' theory, nor the thermal-conduction modification provides an adequate explanation of the experimental facts.

In terms of a proposed *concentration theory*, sensitivity is determined primarily by the concentration of minute receptors. Neglecting a minor effect of size as such, the intensity of experience from stimulation with a stated temperature is proportional to the square root of the number of receptors per unit area, which might be translated neurologically as the square root of the total frequency of nerve impulses per unit area. The question is raised whether the concentration theory may be a fundamental rule for all forms of cutaneous sensitivity, and also whether the square-root function may express a general relation between intensity of experience and frequency of nerve impulses. These are problems for further research.

(Manuscript received May 2, 1939)

THE DEVELOPMENT OF INTRA-SERIAL RELATIONSHIPS IN ROTE LEARNING

BY MAX HERTZMAN AND WALTER S. NEFF

College of the City of New York

The Problem

Although over fifty years, rich in experimentation and analysis, have passed by since the appearance of Ebbinghaus's (1) classic work on learning and retention, many of the basic problems in this field remain unsettled. One of the most fundamental of these incompletely resolved questions is that concerning the nature of the relationships that develop between adjacent and separated items in a serial list, when the list is committed to memory. Since Ebbinghaus's day it has been common psychological practice to speak of the development of associative links or bonds between items to account for the fact that the presentation of one item from a learned list will facilitate recall of other items from that list, although even here the influence of *Gestalttheorie* has made itself felt in an increasing insistence upon organizational principles.[1]

When the learning of serial lists is considered from the viewpoint of traditional associationism, learning is discussed in terms of the development of bonds between items. In connection with this point of view, certain interesting questions develop. Are these bonds formed only between adjacent items (A–B, B–C, C–D, etc.), or are they also formed between remote items (A–C, A–D, etc.)? Are connections formed only in the forward direction (the manner in which the list is generally learned), or do backward associations also develop? What can be said about the relative strength of direct and remote, forward and backward associations? The answers

[1] In this connection, see the discussion of rote learning in Woodworth (7, pp. 23–35).

to these questions, where experimentally obtainable, have been utilized to construct theories and 'laws' of learning.[2]

A number of researchers, using a variety of techniques, have attacked the problem of the nature of intra-serial relationships, with considerable agreement in results.[3] All except a single worker find evidence for the existence of remote associations, and almost all who have worked upon the problem have demonstrated the existence of backward associations. With the exception of these findings, however, the field is still somewhat chaotic. There is no final agreement, for example, on the character of the relationship which holds between the strength of associations and their degree of remoteness, although the general assumption has been (Ebbinghaus, Hull, and others) that the relationship is inverse and regular. Recently some contrary evidence has been presented on this point (see discussion of Raskin and Cook below). Furthermore, the relationships found between forward and backward associations, as well as their significance for theory, have not been made wholly clear. And finally, the part which 'organization' plays in the formation of intra-serial relations, as well as the possibly very important effects of the attitudes of the learner have hardly been treated in the research literature.

The study which provides the immediate background of the present research is that of Raskin and Cook, whose procedure we followed with certain important modifications (see below). Their principle finding was that the curve which presents the relationship between strength of associations and their degree of remoteness is *u*-shaped, *i.e.*, strength of association decreases regularly from immediate direct associations to those of medium remoteness, but then *increases* as regularly until the most remote associations appear much stronger than those of medium order. An analysis of the relationship existing between backward and forward associations in their data convinced them that the fundamental factor making for this

[2] By far the most elaborate set of 'laws' of learning has been developed by Hull (3) on the basis of extant evidence and inferences from this evidence. Hull's 'miniature system' has excited wide attention.

[3] Adequate summaries of the literature in this field may be found in Raskin and Cook (5), and McGeoch (4).

u effect was the influence of circularity, *i.e.*, the fact that *S* will not regard the interval between learning trials as critical but will form low order forward associations *across* the interval. Since terminal syllables, they argue, are therefore psychologically related to the initial syllables (across a 'gap,' to be sure), we have the appearance of quite remote associations being almost as strong as quite immediate ones. This result, they feel, contradicts the basic formulation of Ebbinghaus and Hull on the question of the relationship between direct and remote associations.

A serious shortcoming of all of the experimental studies of this problem so far performed (a defect which they share with the great majority of those in the entire field of rote learning), is that no attempt has been made to study this problem in a genetic (long-sectional) manner. All the investigators here considered have attempted to study the relationship existing between items *which have already been learned.* No attempt has been made to study the *development* of these relationships as learning proceeds toward completion. The typical procedure has been to set *S* the task of learning lists of syllables to a given criterion of 'success' (*e.g.*, three successive errorless reproductions), and then to test for intra-serial relationships by a variety of methods. Such a procedure tends to neglect the fact that rote learning is a dynamic process, and represents a progressive modification of performance on the part of a behaving organism.

The authors of this study, therefore, have set themselves the task of gathering some evidence on the development of intra-serial relationships in an effort to cast some light upon these unsettled problems.

Procedure

Various techniques have been employed in testing the strength and direction of associations, *i.e.*, relearning reversed or derived lists, the study of 'place-skipping' errors, and a controlled association method (Wohlgemuth (6), Raskin and Cook (5), McGeoch (4)). Although it is not free from certain difficulties, the latter technique was the one utilized in the present study.

Materials.—In order to make sure that similar conditions held in our experiment as in that of Raskin and Cook (the clearest in the field), we used the same materials as they employed. These comprise eight nonsense syllables chosen from Hull's (2)

published list. These syllables were printed in large block letters on 4″ by 6″ bristle board cards, so that they could be used conveniently in a group experiment.

Subjects.—The *S*s were students registered in the course in General Psychology at the City College.

Method.—The experiment naturally falls into two parts: (1) the learning series, and (2) the test series. In the first part, the *S*s, who met in groups of approximately 20, were instructed that nonsense syllables were to be presented to them which they were to learn by the method of anticipation, pronouncing the syllables subvocally. They were informed that the series was to be presented until it was evident to the instructor that almost all in the class had learned the list. A very emphatic instruction was laid down that they must keep on learning (anticipating) until the last trial, although they were to note on a piece of paper before them the number of the trial at which they first anticipated all the syllables correctly. By this means an important purpose was served. Since learning efficiency varied from subject to subject, and since a prearranged number of trials was given (12–15), it was possible to obtain examples of several degrees of mastery of the list, ranging from those *S*s who never succeeded in mastering it at all, to those who mastered it very quickly and overlearned the list to the extent of 8 or 9 trials. The time of exposure of each card was 2.5 seconds, the interval between them was minimal, and the interval between trials was about 4 or 5 seconds.

In the test series, the eight original syllables were repeated three times in haphazard order (no syllable was repeated twice in immediate succession). *S*s were instructed that they were to *write*, in response to each stimulus syllable, *the first syllable from the original list that popped into their heads*. In this part of the experiment, *E* read the syllables aloud, allowing about 6 seconds between syllables for the recording of responses. The experiment was over at this point, but before the papers were taken up *E* asked the *S*s to make a record of the following on their data sheets: (1) presence or absence of any deliberate plan of response (nearest syllable, forward association, etc.), (2) presence or absence of grouping in original learning, and type of grouping, if any, and (3) any departure from instructions laid down by *E*.

Analysis of Data.—Because of the degree of complication of the experimental procedure, which not all *S*s could follow, and because of some degree of variation in the attitude which *S* took toward the task, certain papers had to be eliminated. Three classes of papers were discarded: (1) those cases where *S* admitted that he had followed a deliberate plan in giving associations; (2) a few cases in which the instructions had been seriously misunderstood, either in the learning or the test series; (3) two or three cases of relatively complete blocking, where very few responses were given or there was a strong fixation on a single syllable in response. 107 *S*s remained for analysis.

Since our primary purpose was to study the development of any factors involved, this final group of *S*s was divided into four subgroups, as follows. Group I.L. (incomplete learners) comprised 22 persons who had not succeeded in mastering the list in the learning series, but participated in the test series in good faith; Group A consisted of 27 *S*s who had 'overlearned' the list from 1 to 3 trials;[4] Group B contained 32 *S*s

[4] For reasons of terminological convenience, the trial at which mastery first occurs is included in our calculation of degree of 'overlearning.' *E.g.*, if *E* had administered 12 trials to the particular group and an *S* reported his first fully correct anticipation at the end of the 10th trial, he was given credit for 3 'overlearnings' (trials 10, 11, and 12).

who obtained from 4 to 6 overlearnings; Group C contained all those persons (26 in all) who reported from 7 to 9 overlearnings. It is obvious that the divisions are arbitrary, but that the four groups represent, on the whole, four different levels of mastery of the nonsense list, with no implication that the differences between adjacent levels are equivalent.

RESULTS

1. *Strength of the Different Orders of Association*

The frequency of the occurrence of each order and kind of association for each of the four degrees of mastery is presented in Table 1.

TABLE 1

FREQUENCIES AND CORRECTED FREQUENCIES OF THE DIFFERENT ORDERS OF ASSOCIATIONS

Order of Association	Obtained Frequency				Corrected Frequency (in percent)			
	I.L. Group	Group A	Group B	Group C	I.L. Group	Group A	Group B	Group C
First order forward.....	80	134	250	206	17.3	23.6	37.2	37.7
Second order forward....	70	84	85	68	17.7	17.3	14.8	14.5
Third order forward.....	26	42	38	22	7.9	10.4	7.9	5.6
Fourth order forward....	26	22	25	23	9.8	6.7	6.5	7.4
Fifth order forward.....	21	21	16	8	10.6	8.6	5.6	2.9
Sixth order forward.....	9	20	10	16	6.8	12.3	5.2	6.3
Seventh order forward...	4	9	3	3	6.1	11.1	3.1	3.8
First order backward.....	74	66	116	102	16.1	11.6	17.2	18.7
Second order backward...	51	73	72	65	12.9	15.0	12.5	13.9
Third order backward....	27	35	27	23	8.2	8.6	5.6	5.9
Fourth order backward...	15	25	30	20	5.7	7.7	7.8	6.4
Fifth order backward....	26	19	16	9	13.1	7.8	5.6	3.3
Sixth order backward....	10	16	15	10	7.6	9.9	7.8	6.4
Seventh order backward..	10	15	17	26	15.2	18.5	17.7	33.8

* Miscellaneous responses have not been reported.

In order to compare the relative strength of the different orders of association, the obtained frequencies are insufficient, since the number of different ways in which a given association may be obtained is not equivalent for the different orders. Thus a first order association may be obtained in seven different ways, a second order in six different ways, a third order in five different ways, etc. Raskin and Cook corrected for this difference in opportunity of response by multiplying the obtained frequencies by constants which made each one of them comparable to the first order associations. Their method was adequate for their conditions but it cannot be used here since we have to compare groups containing varying numbers of subjects. The corrected values from group to group should be directly comparable. The corrections applied in the present study were obtained in the following manner. The obtained frequency was divided by the maximum number of times that it was possible under the conditions

of the experiment to form the association in question. A first order association, for example, could be formed in seven different ways. Since each stimulus was repeated three times, this means that a single subject could make a maximum of 21 first order associations. The entire group of subjects could make a maximum of 21 × N first order associations, so that the denominator for the response in question would be 21 × N. In a similar manner, we divided by 18 × N for the second order associations, and by 15 × N for the third order, etc.

Raskin and Cook concluded from their data that their subjects learned the series as if it were circular. Instead of each trial being considered as a unit, they inferred from the disproportionate strength of the more remote forward and backward associations that adjacent trials might form a single series. If this were the case, remote associations would in actuality be less remote than they appeared to be on the surface. A sixth order forward association, for example, could also have been a second order backward. All remote associations beyond the fourth order could be seen as an association of opposite direction and of a lower order.

Despite the added strength of the more remote associations, their corrected data (Raskin's and Cook's) exhibit a gradient of decreasing frequency of response for both the forward and backward associations, the gradient being interrupted when the more remote associations are reached. The gradient is as important an aspect of the description of the results as is its inversion which attests to the superior strength of the more remote associations which led to the concept of the circular series.

For the fullest understanding of the associations formed in learning we must have a genetic picture from early parts of the learning process to stages where there has been considerable overlearning. A cross-section of a dynamic process does not give full enough information about the nature of the process. Features of the cross-section may be singled out for emphasis in such a manner as to distort the interpretation of the processes under investigation. A series of cross-sections does not completely eliminate the possibility of this sort of error but it provides for study at different stages thus indicating, at any rate, a sequential relationship. Different phases of a developmental process are less likely to be singled out as being central

if the segmentation makes it clear that changes are occurring. The role that a given phase is playing may then be more correctly understood.

The data which enable Raskin and Cook to emphasize the concept of circularity represented only a cross-section. The frequencies of associations which they found are characteristic of the situation where there has been a small amount of overlearning.[5] With the four degrees of mastery utilized in the present experiment we were more nearly able to get a picture of the development of the associations than a single cross-section could give us.

In Table 1, the corrected percentages for each degree of mastery and for each kind of association are presented. We find that the percentages for the incomplete learning, representing our lowest degree of mastery, do not give us a very consistent picture. Among the forward associations there is some indication of the beginning of a response gradient but with an inversion at the fifth order association. While this inversion would be consistent with the data of Raskin and Cook, the tendency for circularity which it might indicate is not supported by the frequencies of the sixth and seventh orders, which are lower than the third and fourth. The more remote backward associations do not consistently increase in frequency. The association between the first and last syllables is relatively as strong as in the data reported by Raskin and Cook. In general, we may say that the data for our incomplete learning group is relatively unpatterned and indicates the probable operation of a variety of tendencies whose nature cannot be clarified until we go further into the developmental process.

The degree of mastery of the individuals in our A group is similar to that permitted for all the subjects by Raskin and Cook. The data obtained is also similar to theirs, with the more remote associations among both the forward and backward appearing in increasing frequency. With respect to the forward associations, we have an increase in the relative proportion of first order associations.

[5] Their criterion of learning was three successive correct repetitions.

For our next highest degree of mastery (B), a very definite gradient has been established with respect to the forward associations. The backward associations exhibit a sharpening of the gradient in frequency of responses as the associations become more remote. There is no specific pattern but the data are unlike the corresponding data for Group A.

For our highest degree of mastery, Group C, the most marked deviations from Group B lie in the large increase of seventh order backward associations which may be thought of as being first order forward associations if there is some circularity in the learning of the series. The frequencies do not produce a very regular series, but beyond the second order the associations are very small in number. The main characteristics appear to be a heightening of the strength of low order associations with a certain irregularity of the frequencies for the remote orders.

2. *The Relative Frequencies of Forward and Backward Associations and the Development of Gradients*

The discussion of the general patterns of response shown in our four groups would indicate the necessity of analyzing the relation between our two kinds of responses as the mastery of the test increases, as well as the development of differential frequencies as the associations become remote. In Table 2, the relative frequency of forward and backward associations for our different degrees of mastery, and some ratios between the frequencies of low order associations and more remote orders, are presented.

a. Forward and Backward Association

As the mastery of the test becomes more complete there is a tendency for the frequency of the forward responses to increase. This tendency is more marked if the seventh order backward associations are treated as first order forwards. These results would indicate that the method of learning is influencing the direction in which the associations are being formed.

b. The Development of Gradients

As the mastery of the test becomes more complete, the remote associations become relatively less frequent. The development of this gradient is not as marked for the backward associations as it is for the forward ones. Thus we find an increase of from 33.9 to 59.5 percent in the frequency of first order forward associations as we go from our lowest to our highest degree of mastery; for the backward associations, on the other hand, the changes are not very marked, the frequency of first order association going from 34.7 to 40.0 percent with an inversion in Group A.

The development of demarcations between the first order associations and the more remote ones can be further seen when the ratio between first and second order associations are compared. By dividing the corrected percentage of first order associations by the corrected percentage for the second order, we find a rise from a ratio of practically one for the incomplete learning to a ratio of 2.6 for group C (see line 10, Table 2).

TABLE 2

RATIO OF SELECTED FORWARD AND BACKWARD ASSOCIATIONS TO THE TOTAL NUMBER OF ASSOCIATIONS (IN PERCENT)

Nature of the Ratio	I.L. Group	Group A	Group B	Group C
1. Between total for. and all associations	52.6	57.1	59.3	57.6
2. Between total back. and all associations	47.4	42.9	40.7	42.4
3. Between first order for. and total for.	33.9	40.4	58.5	59.5
4. Between first order back. and total back.	34.7	26.5	39.6	40.0
5. Between second order for. and total for.	29.7	25.3	19.9	19.7
6. Between second order back. and total back.	23.9	29.3	24.6	25.5
7. Between first and sec. order for. and total for.	63.6	65.7	78.5	79.2
8. Between first and sec. order back. and total back.	58.6	55.8	64.2	65.5
9. Between seventh order back. and total back.	4.7	6.0	5.8	10.2
10. Between corrected percent of first order and sec. order for.	97.7	136.4	251.4	260.0
11. Between corrected percent of first order and sec. order back.	124.8	77.3	137.6	134.5

For the backward associations, the rise is very slight and is not steady. Other ratios indicating the sharpening of the gradient for the forward associations are presented in Table 2. These results would indicate that as the amount of learning increased there was an increase in the frequency of associations

in a way that corresponded to the manner in which the learning was accomplished, as shown by the increase in first order responses in the forward direction without an increase of a comparable nature of the low ordered backward associations. These results would point to a decrease of the influence of circularity with an increased mastery.

3. *The Factor of Familiarity*

While the data do not clearly show that preference in utilizing syllables that occur in certain serial positions rather than others produces the patterning of associations as shown in Table 1, it is clear that the relative frequency with which a syllable is used as a response is altered with an increased degree of mastery, the tendency becoming one in which all syllables are given with equal frequency as mastery of the series increases. Differential familiarity may affect the patterning of associations, since of the fourteen different associations possible in an eight syllable list, each syllable can be at the response end of only seven different ones. To take an extreme case, the syllable in the first position can be used as the response only in a backward association. The large number of backward associations to be found in the early stages of mastery might be partly due to this factor. For the incomplete learning, the last three syllables of the list are given most frequently as responses; as learning progresses, the relative frequency with which each syllable is used becomes more uniform, as is shown by the mean deviation of these frequencies from expectancy.[6] The mean deviations going from the incomplete learning group to Group C are 4.51, 1.95, 1.86, and 1.39 respectively.

Discussion

As research piles up in many fields of learning, it has become increasingly evident that certain plausible and authoritative explanations and theories have suffered from oversimplification. The proponents of one theory or another have too often achieved striking formulations by seizing upon only

[6] By chance each syllable would be used 12.5 percent of the time.

one aspect of an extremely complex phenomenon. Those who have approached our problem from the point of view that learning consists in forming associations have preoccupied themselves almost exclusively with a study of the conditions of formation of relationships between items, and have attempted to formulate 'laws' about these relationships.

Certain consequences flow from this theoretical position in the light of which our data must be examined. According to association theory, and under strict operation of the law of contiguity, we should expect: (1) a direct relation between remoteness of syllable and frequency of response, and (2) that the *relative* strength of the different orders of association remain relatively the same throughout the course of learning.

It should be clear that the factor of contiguity is not sufficient to explain our results. These results may be due to the influence of several factors, some of them 'attitudinal' in character, which operate during the learning process. It is possible, further, that these factors operate with differential effects at different stages of the learning process. Our experimental results do not point unequivocally to one or the other process as operating predominently throughout. It must be realized that our learner is a living and complex organism, subject to the changing influence of such intangibles as attitude toward the task, mode of attack, self-instruction, degree of acceptance of instructions, and so on.

Factors that probably influenced the kind and frequency of responses were the following: (1) The mode of presentation, which could make both for (*a*) forward association and (*b*) a heightened frequency of remote backward associations (really low order forwards) because of the influence of circularity; (2) Differential availability of the nonsense syllables as responses during different stages of the learning; (3) A variety of attitudes which could be developed and altered during the course of learning and which could affect the mode of response. Such attitudes might include (*a*) attempts to give forward associations because of the mode of presentation, (*b*) when a high degree of mastery had been achieved, the resulting tendency to give first order forward associations as the easiest

to ellicit, or (*c*) the tendency to give responses at random from the list because of equal ease in manipulating all the elements of the list.

We find that for our lowest degree of mastery (incomplete learning), there is no particular response pattern, but the more remote associations (fourth order and greater) are at least as frequent in occurrence as they are in the higher degrees of mastery. This situation may not be due to association in the causal sense (*cf.* Woodworth, **7**, p. 25), but is a descriptive fact whose existence may be due to the differential availability of syllables as responses (primacy-finality effect).

The first general pattern that develops (Group A) is that which was demonstrated by Raskin and Cook, *i.e.*, a u-shaped pattern where the medium remote associations are given less frequently than either the low or high order associations. This appears to be a consequence of both a tendency for forward associations and one for circularity. It should be noted that as mastery of the task is increased (through the higher degrees of mastery), the forward tendency strengthens and the effect of circularity appears to become less prominent, the u-effect tending to disappear. This phenomenon indicates that forward direction and circularity need not go together. As the material becomes more consolidated there is an increased heightening of the strength of the more immediate forward associations; this effect may be a function of the mode of presentation of the material. This phenomenon is not so clear-cut for the backward associations. In general, increased mastery leads to what appears to be a sharpening of the association gradient. The slope of the curve which shows the relationship between degree of remoteness and frequency of response becomes steeper. This is shown further by the marked increase of seventh order backward associations which, with even a limited circularity, can only be understood as first order forwards.

We have not been able directly to determine the influence of attitude on responses after a considerable amount of overlearning. The data for our group with the highest degree of mastery (Group C) do not indicate a regular pattern other

than the continued high value of the first order associations. At this stage of development *S* may be at liberty to select responses randomly or give automatic first order associations. The presence in the group of *S*s who performed predominently in one fashion or another could give rise to the irregularity observed in our data, but we have no conclusive evidence that this is actually the case.

(Manuscript received March 22, 1939)

References

1. Ebbinghaus, H., *Memory*, Teachers College, Columbia Univ., N. Y., 1913.
2. Hull, C. L., The influence of caffeine and other factors on the phenomena of rote learning, *J. gen. Psychol.*, 1935, 249–274.
3. ——, The conflicting psychologies of learning—A way out, *Psychol. Rev.*, 1935, **42**, 491–516.
4. McGeoch, J. A., The direction and extent of intra-serial associations at recall, *Amer. J. Psychol.*, 1936, **48**, 221–245.
5. Raskin, E. and Cook, S. W., The strength and direction of associations formed in the learning of nonsense syllables, *J. exp. Psychol.*, 1937, **20**, 381–395.
6. Wohlgemuth, A., On memory and the direction of association, *Brit. J. Psychol.*, 1913, **5**, 446–465.
7. Woodworth, R. S., *Experimental Psychology*, Henry Holt & Co., N. Y., 1938.

STUDIES IN HIGH SPEED CONTINUOUS WORK. III. INITIAL SPURT AND WARMING UP

BY B. R. PHILIP

Queen's University

Meaning of the Terms

The definition of initial spurt, as given by Thorndike (1), implies that "at the very beginning a person tends, other things being equal, to work at a higher efficiency than ever again." The concept is none too clear. We do have a guide, given by both Kraepelin (2) and Thorndike (1), that the initial spurt, if a real fact, will be found in an examination of the work, minute by minute for the first quarter of an hour, and we are warned to place careful consideration of the unreliability of any determination based on a few days' record. Furthermore we have Chapman and Nolan's caution that the weakness of previous experiments on initial spurt has been the use of successive intervals of time of too long duration to record the rapid decline in efficiency which takes place, and that consequently the effect of the spurt has been so neutralized as to escape attention.

Obviously the work must be carried on for a sufficient period to enable one to gauge the fluctuations with some degree of accuracy; also a sufficient time must have elapsed since the last work period to ensure that *S* does not start with any perceptible degree of fatigue, otherwise, even if there should be an initial increase in efficiency, it might be an effect of warming up. Finally we should not always expect the period of maximum efficiency to be at the very outset. Initial adaptation, concentrating of one's attention, controlling of emotional factors, and even the rapid improvement of practice, may postpone the point of maxium efficiency in such a simple operation as tapping for several hours after the beginning of work, when very evident fatigue must have cut down the efficiency to a great degree.

In general what is recognized as initial spurt, is the attainment close to the beginning of work of a high efficiency which is followed by a substantial drop. So highly motivated is *S* that he strikes a working rate, which, in his present state of practice, he cannot long maintain. The chief clue to the initial spurt is not the maximum efficiency at the very outset, but rather the presence of its subsequent pronounced decrement. After this decrement, which generally carries the efficiency below the effective maximum spoken of in a previous paper (**4**), *S* may and nearly always does improve his working rate, in spite of fatigue and boredom.

Warming up is a phenomenon of the work curve that occurs towards the outset of the resumption of work when only a short interval of time has elapsed since the last work period. The time interval is of great importance. If the interruption is very short, so that there is practically no break in the continuity of the work, the efficiency continues undiminished, except for the random fluctuations which always occur, and the periodicity which is a definite characteristic of long continued work of this sort. On the other hand if the rest period is so long that *S* completely recovers from fatigue, we are dealing with another initial spurt.

The effect of the discontinuity of the work, emphasized by Robinson and Heron (**6**), was originally pointed out by Wells (**7**). The difficulty of determining the warming up effect quantitatively seems to consist in separating it from the practice effect. Robinson and Heron attempted to devise a formula for so doing, but they themselves have been forced to fall back on the mere inspection of the curves of work; for as Skaggs (**8**) has pointed out, the formula may hide actual (especially small), warming up effects. Possibly the most important drawback to the formula is the fact that it is dependent upon the resumption of the work at a later period to determine the effect of practice. Where the work is continued for some length of time as in the present experiment, the practice effect and random fluctuations may be estimated with some degree of assurance, so that the warming up may be shown directly as a deviation from the normal course of the curve.

The definition of warming up is very diffuse. As stated by Thorndike (1), "for the first half-hour or so, other things being equal, efficiency increases gradually. This we may call the doctrine of 'Incitement' or 'Warming-up effect.'" But later he defines it more precisely as "that part of an increase of efficiency during the first 20 minutes (or some other assigned early portion of a work period) which is abolished by a moderate rest, say of 60 minutes." Skaggs (**8**) uses the term warming up to signify 'improvement in any performance or increase in efficiency over and above the improvement due to practice or actual learning,' and is inclined to the view that "the warming up phenomenon is largely a matter of attitude, initial alertness and attention." He states also that "in any case where muscular adjustments or co-ordinations are involved it is probable that one must also take into consideration a warming up process in the muscle tissue itself." Robinson and Heron (**6**) define warming up as "a rise in efficiency which is steeper and more temporary than the rise which can be seen in, let us say, successive daily performances."

The popular conception of warming up when applied to a task that involves the grosser musculatures of the body means the period of fore-exercise that is required to bring the individual to a high pitch of efficiency. When the muscles have become stiffened from fatigue, or from disuse in the particular function involved, as in the case for instance, of the baseball pitcher before the game, or of the sprinter before the start of the race, a preliminary period is taken during which the athlete exercises with increasing energy till he is able to perform a function with maximum speed and efficiency. Naturally he fatigues himself during the process, but the loss in efficiency due to the fatigue is small compared to the gain from the warming up. The more fatigued a muscle, and the stiffer it has become during a rest, the more pronounced the gain from the warming up, provided exhaustion has not set in.

The popular concept applied in some degree to the present experiment. The warming up effect refers to the rise in efficiency in the period of work following a short interruption, during which *S* does not completely recover from fatigue, be-

fore he reaches his effective maximum working rate, which he has attained or exceeded previous to the rest period. It takes into account: (*a*) attention adjustments, (*b*) the increasing effectiveness of the muscle which has become stiffened, and (*c*) the acquiring of a smooth working rhythm. It is to be noted that there is evidence, when intervals are taken sufficiently short, of a rapid gain not due to practice, at the very outset of the work period, before the point of maximum effectiveness is reached. For this reason the time interval between work periods is of importance in differentiating between the initial spurt and the warming up effect.

Of importance, too, is the fact that the warming up is not followed by a pronounced slump such as occurs with the initial spurt. The latter is characterized by work done above the effective maximum working rate, which it is impossible to maintain, not so much on account of dwindling motivation, as of blocking and cramping; subsequent loss of rhythm and effectiveness brings the rate well below the effective maximum. The former, the warming up, is characterized by work done below the effective maximum working rate, due in part to attention difficulties, but chiefly to the stiffening of the muscle, so that the rhythm is not readily regained. Hence it is that the warming up is not followed by a pronounced slump. A decrease in efficiency ensues, but it follows the course of the work decrement in general.

Experimental Results

Two experiments on high speed continuous work, previously studied for periodicity and decrement, afford data for a comparison of the initial spurt and warming up effects. The description of the apparatus and the general experimental technique have been reported and will not be repeated here.

Figure 1 graphs the scores min. by min. of the 12 *S*s of Experiment A. To study the initial spurt, on one side of the graph are the average scores of the 10 min. preëxperimental test which took place a week before the main experiment, and paired with them are the average min. scores for 10 mins. at the beginning of the main test. For comparison sake, to show

the warming up effect, the 10 mins. preceding a rest or an interruption are averaged, and paired with the 10 min. following the rest. No such rest or interruption occurred for 2 *Ss*, but there were 31 such rests for the remaining 10 *Ss*, *i.e.*, from 1–5 rests per *S*, and they averaged 8.4 mins.

From the figure it may be seen that there is for the preëxperimental test, a small initial spurt, a short slump at the 4 min. mark, a rather irregular course, with the end spurt starting at the 8 min. mark. The last min. was slightly less in efficiency than the 9th. min. as *Ss* had started their spurt

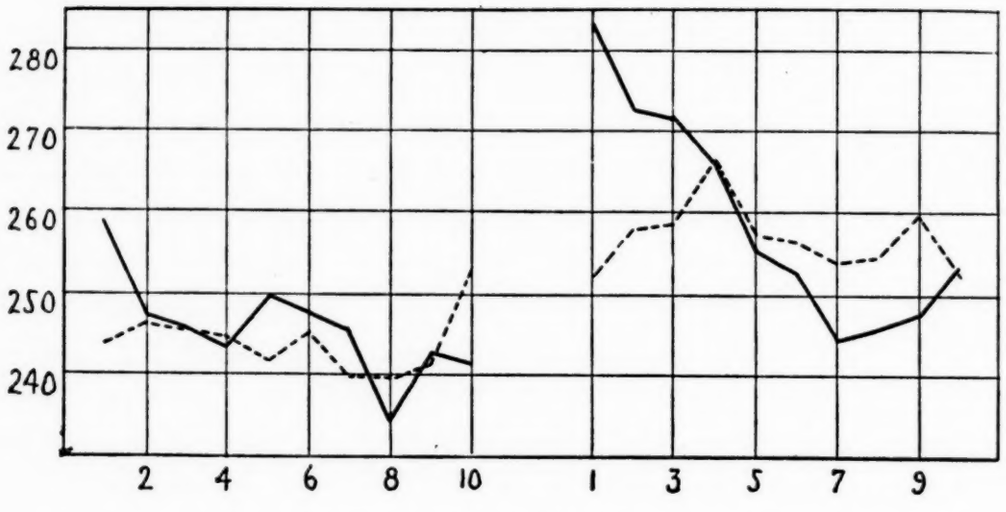

FIG. 1. Comparison of the Initial Spurt with the Warming Up. The continuous line represents the Initial Spurt; the curve on the left is the preëxperimental test; the curve on the right is the beginning of the long work period ($N = 12$). The dotted line represents the Warming Up; the curve on the left is the 10 min. period just before the interruption; the curve on the right, that of the work just after the interruption ($N = 31$). Ordinates, taps per min. Abscissæ, mins.

too soon. As a matter of fact only 4 of the 10 *Ss* had a higher score on the 9th. min. than they had on the last min., but the averaging of the results tends to cut down the variability, and to mask the intensity of the effects. This applies a fortiori to the warming up, due to the maxima coming at slightly different times. Since the phenomena are plainly evident from the average curves, they must be recognized to occur with greater intensity in the individual work curves.

The last 10 mins. prior to the interruptions are of uniform efficiency, with a definite end spurt starting at the 8th. min.

In 10 of the 31 cases, Ss had no knowledge of the approaching interruption; hence there was no end spurt, and consequently the spurts, when they did occur, were more intense than is here represented.

The second half of the curves brings out vividly the distinction between the initial spurt and the warming up effects. It is to be noted that the initial spurt starts immediately (That is when min. readings are averaged; subsequently this will be shown not to hold when 10 sec. readings are taken). There is an abrupt slump lasting till the 7th. min. when the normal course of the curve begins. The initial spurt starts well above the effective maximum, below which it drops during the slump; only about the 10th. min. is this effective maximum regained and the practice effect starts to manifest itself. On the other hand the graph of the warming up shows little or no variation from the effective maximum, except a steady rise from the first to the 4th. min. from which point the regular decrement starts. Most rests took place late in the work period when *S* was really fatigued, and the decrement curve had been fairly well flattened out. During the rest there is, on the average, no loss or gain in effectiveness—*S* starts again precisely where he left off. By comparison the initial spurt curve shows a considerable gain from the end of the preëxperimental test on week previous; most of this gain is lost in the first few mins.

Time of Occurrence of the Maximum

A factor of importance in the study of the initial spurt and the warming up effects is the time of occurrence of the maximum. As has been stated, even at the beginning of the work in which he has been thoroughly practised, *S* generally requires some time to reach his maximum. In a working period which lasts several hours, the highest peak does not usually occur till long after the start. In the case of 12 Ss who performed Experiment A the average time of this occurrence was 146 mins. and only one *S* reached his maximum in the first min. when min. intervals were used. When 5 min. intervals were used 4 Ss reached their maximum in the first interval,

and the average time of occurrence for the group was 80 mins. The 2 Ss in Experiment B, (5), likewise showed a delay in the occurrence of their maximum when 10 sec. intervals were used.

Table I shows the location of the high points for intervals of different lengths, and carried over different times for Ss of both Experiments A and B. As the intervals get smaller it is apparent that, even for the initial spurt, only 2 Ss in Experiment A worked at their immediate maximum in the first 10 secs. and yet they did not reach their min. and 5 min. maximum until after from 2–5 hours. The point to be empha-

TABLE I

FREQUENCY OF LOCATION OF HIGHS FROM BEGINNING

Exper. A	Interval	N	1	2	3	4	5	6	7	8	9	10	11	12	Aver.
Initial Spurt	5 mins.	12	5	1			1	1	1	1			2		4.60
	1 min.	12	6	2	1	2						1			2.60
	10 secs.	10	2	1	4	1			7				1		3.80
Warm Up	5 mins.	31	16	4	2	5	1	1	1	1					2.24
	1 min.	31	2	3	5	8	7	4	1	1					4.16
	20 secs.	26		3			1	2	4		3	2	4	7	9.35
Pre-test	1 min.	12	6	1			1	2	1	1					3.33
Exper. B															
S1	1 min.	26	23	2	1										1.15
	10 secs.	26	6	9	1	7	1		1				1		3.00
S2	1 min.	28	28												1.00
	10 secs.	28	5	22		1									1.89

sized is that even for the initial spurt the time of maximum working rate, when the intervals are sufficiently fine, does not usually occur at the outset. In the case of the warming up, *i.e.*, when an interruption has taken place, the occurrence of the maximum point is definitely delayed. Time must be taken to achieve maximum working speed when S is free from fatigue; when fatigue is present and the muscles have become stiffened after rest, it takes longer to bring about the effective maximum.

RELATIONSHIP BETWEEN VARIOUS POINTS IN THE CURVE

Of some interest is the relationship between the high and the low points, and the averages over different work periods.

As there is quite a variation in the working rate of different Ss, naturally these relationships should be expressed as ratios, preferably percentages. These are shown in Table II*A*. The

TABLE II

(*a*) 10 min. aver. before rest; (*b*) Starting speed after the rest; (*c*) 10 min. aver. after rest; (*d*) Low rate; (*e*) High rate; (*f*) Hour rate.

A. Mean Ratios in Percentages

		Initial Spurt						Warming Up					
2nd Term		(*a*)	(*b*)	(*c*)	(*d*)	(*e*)		(*a*)	(*b*)	(*c*)	(*d*)	(*e*)
	(*a*)	—					(*a*)	—				
	(*b*)	116					(*b*)	101				
First	(*c*)	107	92				(*c*)	105	102			
Term	(*d*)	93	81	89			(*d*)	94	91	88		
	(*e*)	119	104	115	128		(*e*)	116	113	110	122	
	(*f*)	107	93	99	115	87	(*f*)	102	99	103	109	87

B. Reliability of the Differences (Mean/P.E. Mean)

	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)
(*a*)	—				
(*b*)	4.4				
(*c*)	1.0	5.2			
(*d*)	0.3	5.9	0.7		
(*e*)	1.2	4.8	2.9	3.4	
(*f*)	2.4	2.9	1.8	3.0	0.5

C. Intercorrelations

	Initial Spurt ($n=12$)						Warming Up ($n=31$)				
	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)		(*a*)	(*b*)	(*c*)	(*d*)	(*e*)
(*a*)	—						—				
(*b*)	.46						.79				
(*c*)	.59	.72					.83	.92			
(*d*)	.63	.50	.90				.78	.94	.95		
(*e*)	.34	.57	.79	.57			.80	.88	.96	.93	
(*f*)	.67	.11	.47	.52	.27		.91	.81	.90	.87	.90
	P.E. From 0.04 to 0.20						P.E. From 0.02 to 0.05				

relationships were first worked out for each individual *S* in Experiment A, and then averaged; but this involves unnecessarily tedious computing, for the ratios of the averages yield the same results within about 1 percent. To gauge reliabili-

ties of differences, standard deviations were obtained from ratios calculated by the first method. Based on 1 min. intervals 13 different points were selected on the curves of initial spurt and warming up. Of these 6 were finally selected, for which all ratios were computed.

The points chosen were as follows:

(*a*) Curve I, the 10 min. average of the preëxperimental period; Curve II, the 10 min. average just prior to the rest period. The remaining points refer to the rate per min. of the work done at the outset of the main test for Curve I, and immediately after each interruption for Curve II.
(*b*) The initial min. working rate.
(*c*) The initial 10 min. average.
(*d*) The low point of the first 10 mins.
(*e*) The high point of the first 10 mins.
(*f*) The average rate per min. of the first hour.

To bring out differences between the curves of the initial spurt and the warming up, Table II*A* presents the ratios in the two curves side by side, and Table II*B*, the reliability of the differences between these ratios. It is seen that practically all the ratios involving (*b*), the starting speed, are reliable. This is what we should expect, for the initial spurt is characterized by a high starting speed with a subsequent slump below the effective maximum working rate; while the warming up has its maximum later in the curve. A point of some interest is the low probable error of the ratio between the low point and the 10 min. average after the rest—the low point being for both curves about 88 percent of the 10 min. average. The data consistently show a much greater uniformity among the low points than among the high points, and prediction based upon them is more reliable.

Table II*C* presents intercorrelations based on the same variables for both curves. The correlations were obtained by the Spearman Rank Difference method and the rho's were transmuted to Pearson *r*'s. The correlations for the warming up are higher than for the initial spurts, averaging 0.88 as

compared with 0.51. Naturally one would expect the time factor to be one of the chief causes of the higher correlations for the warming up, for the intervals between the two parts of that curve averaged about 8 mins., while the preëxperimental work was done one week previous to the main test. But this would hold only for the ratios involving (*a*); and the average value of these correlations is about the same as those of the whole group. The explanation probably lies in the greater variability at the outset of the work, before *S* has steadied away. The high average correlations for the low point (*d*) imply more uniformity from *S* to *S* for this point.

Effect of the Size of the Interval

We have previously emphasized the role played by the size of the interval used in determining the initial spurt and the warming up. Figure 2 shows the relationship in Experiment A between the form of the curves and the varying size of the

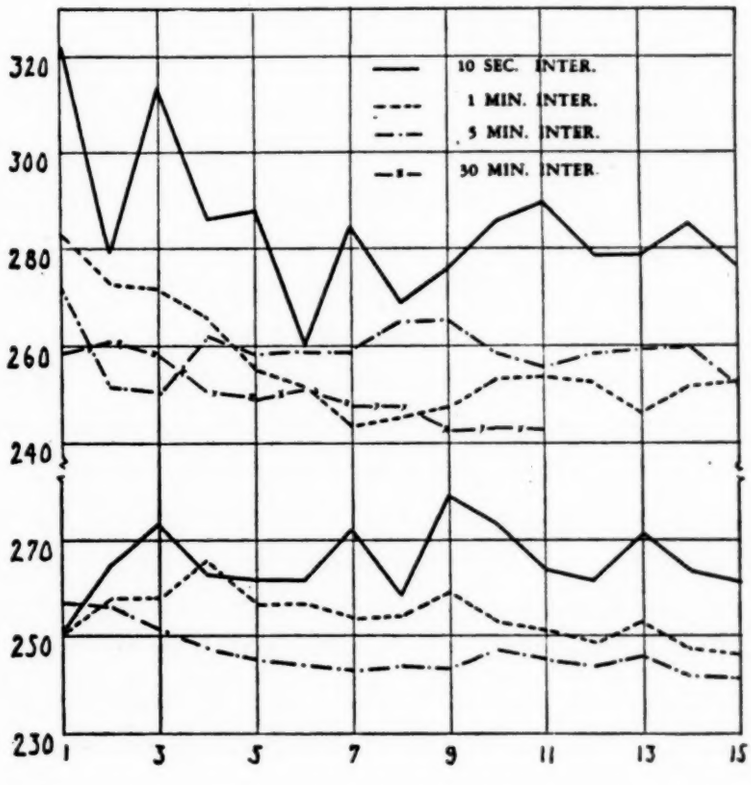

Fig. 2. Effect of size of interval on the Initial Spurt and Warming Up. Initial Spurt, upper curves; Warming Up, lower curves. Ordinates, taps per min. Abscissæ, mins.

interval. The intervals chosen were: 10 sec., min., 5 min., and 30 mins., which increase almost logarithmically. The effects are studied in most cases for 15 intervals. Only 11 of the longest intervals are used for the initial spurt. As the interruptions occur mostly towards the end of the working period, there was not a sufficient number of 30 min. intervals which could be used to show the relationship in the warming up.

To be noted is the similarity of the curve forms no matter what the size of the interval. Thus in the case of the initial spurt the 10 sec. and the 5 min. intervals both show an early slump followed by a subsequent rise. The 30 min. interval is of course too long to show any initial spurt, but it does indicate the high point of the curve to be at the end of the first hour, after which the decrement is slow but continuous.

There is a definite similarity between the curves for warming up, estimated on the 10 sec. and min. intervals;—a short abrupt rise, a slight slump, and a second rise, followed by a slower slump. The warming up does not show in the 5 min. curve. In general the initial spurt curves show greater variability than those of warming up.

The decrement, even after 5 or 6 hours of work, is surprisingly small. From the high point, which occurs at the end of 60 mins., to the low point, which occurs at the $4\frac{1}{2}$ hour mark, there is only a drop of 17.6 taps per min.—a decrement of 6.7 percent. The work curves of 3 *Ss* went to the 14th interval, which would increase their decrement. Even when the high and the low 30 min. periods of these *Ss* are compared, the latter is on the average 84 percent of the former. Yet this does not give us a fair picture of the decrement, for only 6 of the 12 *Ss* worked at their lowest efficiency in the last 30 min. period; and on the average, the low point was reached 2.7 periods before the end of the work, so that the real average decrement must be still less than the 16 percent aforementioned. In spite of the length of time the work was continued *Ss* were far from exhaustion.

The presentation of the results in the form of averages diminishes the amounts of the decrement, as estimated by the

ratio of the low to the high points. From Table III it may be seen that, depending on the size of the interval chosen, the decrement is increased considerably when calculated for each individual *S*, rather than from the average of the group. In

TABLE III

RATIO OF LOWS TO HIGHS IN THE FIRST INTERVALS

Exper. A	Interval	N	Time in Mins.	Aver. of Ratios	Ratio of Averages
Initial Spurt	10 secs.	12	2½	69	81
	1 min.	12	10	77	86
	5 mins.	12	75	83	93
	30 mins.	12	330	84	91
Warm Up	10 secs.	26	2½	62	90
	1 min.	31	10	80	94
	5 mins.	31	75	86	94
Exper. B					
S1	10 secs.	26	2	72	84
	1 min.	26	10	80	83
S2	10 secs.	28	2	82	88
	1 min.	28	10	79	81

general, the smaller the interval chosen, the greater the relative decrement; and the decrement of the initial spurt is more pronounced than that of the slump which follows the warming up.

(Manuscript received January 26, 1939)

REFERENCES

1. THORNDIKE, E. L., *Educational psychology*, New York, 1913, III, Ch. III.
2. KRAEPELIN, E., Die Arbeits Curve, *Philos. Stud.*, 1902, **19**, 451–502.
3. CHAPMAN, J. C. AND NOLAN, W. J., Initial spurt in a simple mental function, *Amer. J. Psych.*, 1916, **27**, 256–262.
4. PHILIP, B. R., Studies in high speed continuous work. I. Periodicity, *J. exp. Psych.*, 1939, **24**, 499–510.
5. PHILIP, B. R., Studies in high speed continuous work. II. Decrement, *J. exp. Psych.*, 1939, **25**, 307–315.
6. ROBINSON, E. S. AND HERON, W. T., The warming up effect, *J. exp. Psych.*, 1924, **7**, 81–97.
7. WELLS, F. L., Normal performance in the tapping test, *Amer. J. Psych.*, 1908, **19**, 437–483.
8. SKAGGS, E. B., A study of warming up in the case of a task of more complicated perceptual motor co-ordination, *J. Appl. Psych.*, 1931, **15**, 499–511.

DISCUSSION

OPTIMAL ORDERS IN THE METHOD OF PAIRED COMPARISONS

BY ROBERT T. ROSS

Stanford University

In a recent discussion of orders for the presentation of items in the method of paired comparisons, Wherry (2) criticizes orders earlier prepared on a rational basis by Ross (1). A further statement appears to be desirable in order to clear up certain misinterpretations of Ross's proposals and some misunderstandings by Wherry of the nature and meaning of his own.

I. Wherry first objects to the Ross lists on the grounds that they are not optimal with respect to 'space errors.' He has no improvement to offer with respect to 'time errors' (that is, the separation of pairs involving the same items).

With regard to space errors (that is, whether an item is presented equally often on the right and on the left), a somewhat different situation exists. The criterion of maximal spacing ('time errors') is entirely unconnected with the order of presentation *within* any pair. In 1934 Ross wrote, "The series obtained from the formulæ presented may be 'balanced'; that is, so arranged that any given number appears an equal number of times as the first and second member of a pair." The orders presented by Ross fulfill this criterion of 'balance.'

What Wherry objects to is that any item does not *alternate* from first to second position in subsequent pairs as one reads through the list. There was no attempt in Ross's lists to satisfy this additional criterion. If conformance to this criterion is required, the Ross lists may easily be so rearranged. In Table I, columns B and E, are the Ross lists for 5 and 7 items as published in 1934 with the 'space errors' indicated by asterisks. These are the 'errors' demonstrated by Wherry. In columns C and F the same lists are rearranged so that they have only 2 and 4 such errors respectively, and it will be shown in section III of this paper that this arrangement is 'optimum.'

What Wherry has said in effect, therefore, is that the Ross lists were not arranged according to the criterion which Wherry has imposed. There is no argument about this. Wherry has assumed that the Ross lists cannot be so arranged. This was a mistake, as has been empirically demonstrated above, and will be rationally demonstrated below.

TABLE I

VARIOUS ARRANGEMENTS OF THE ROSS LISTS

A	B	C	D	E	F
1–2	1–2	1–2	1–2	1–2	1–2
5–3	5–3	3–5	7–3	7–3	3–7
*1–4	4–1	4–1	6–4	6–4	4–6
2–3*	3–2*	2–3	*1–5	5–1	5–1
5–4	*4–5	5–4	2–3*	3–2*	2–3
1–3	1–3*	1–3*	*7–4*	4–7	7–4
2–4	2–4	4–2	*6–5*	*5–6	6–5
*1–5	5–1	*5–1	*1–3*	1–3	1–3*
3–4*	3–4*	3–4	*2–4*	2–4	4–2
2–5	*2–5	2–5	*7–5*	7–5	5–7
			*1–6	6–1	*6–1
			3–4*	4–3*	3–4
			2–5	5–2	2–5
			7–6	*6–7	7–6
			1–4	1–4	1–4*
			3–5	3–5	5–3
			2–6	2–6	6–2
			*1–7	7–1	*7–1
			4–5*	4–5*	4–5
			3–6	*3–6*	3–6
			2–7	*2–7	2–7

A and D are lists as read directly off the 1934 Ross matrix.

B and E are the same lists arranged according to Ross's criterion for balance. These lists were published in 1934.

C and F are the same lists arranged to conform to Wherry's additional criterion of 'space' balance.

* denotes a 'space' error according to the Wherry criterion.

II. Wherry's next step is to derive, 'after a certain amount of trial-and-error,' a list which he maintains is superior to the Ross list. He discovers that as far as 'time errors' are concerned in his list "no further manipulation succeeded in bettering this result; many other schemes were worse"; but that where 'space errors' are concerned, his list contains 'just one half of the incidence of this type of error in the Ross optimum list.'

As has been shown above, the Ross list may be rearranged to give this same result. In addition to this, if in Wherry's example

for seven items we substitute 5, 7, 4, 2, 1, 3, 6 for 1, 2, 3, 4, 5, 6, 7, respectively, and turn the resulting list upside down, we discover that Wherry empirically derived the Ross list for seven items, but failed to recognize it. Similarly for his list of five items, if we substitute 5, 2, 4, 3, 1 for 1, 2, 3, 4, 5 in the Wherry list and invert it, we obtain the Ross list for five items, as shown in Table II.

TABLE II

THE EQUIVALENCE OF ROSS'S RATIONAL LISTS AND WHERRY'S EMPIRICAL ONES

A	B	C	D	E	F	G	H
1–2	5–2	2–1	1–2	4–2	2–7	1–2	1–2
3–4	4–3	5–3	5–3	6–7	3–6	3–7	7–3
5–1	1–5	1–4	4–1	3–1	4–5	4–6	6–4
2–3	2–4	3–2	3–2	2–5	7–1	5–1	5–1
4–5	3–1	4–5	4–5	7–4	6–2	2–3	3–2
3–1	4–5	3–1	1–3	1–6	5–3	7–4	4–7
4–2	3–2	2–4	2–4	5–3	1–4	6–5	5–6
5–3	1–4	1–5	5–1	2–7	7–6	1–3	1–3
1–4	5–3	4–3	3–4	4–1	2–5	4–2	2–4
2–5	2–1	5–2	2–5	6–3	3–4	5–7	7–5
				7–5	6–1	6–1	6–1
				1–2	5–7	3–4	4–3
				3–4	4–2	2–5	5–2
				5–6	1–3	7–6	6–7
				7–1	6–5	1–4	1–4
				2–3	7–4	5–3	3–5
				4–6	2–3	6–2	2–6
				1–5	5–1	7–1	7–1
				3–7	4–6	4–5	4–5
				6–2	3–7	3–6	3–6
				5–4	1–2	2–7	2–7

A and E are the Wherry empirical lists.
B is list A after substituting 5, 2, 4, 3, 1 for 1, 2, 3, 4, 5 respectively in A.
F is list E after substituting 5, 7, 4, 2, 1, 3, 6, for 1, 2, 3, 4, 5, 6, 7 respectively in E.
C and G are B and F inverted.
D and H are the Ross lists published in 1934.

Again, Wherry failed to realize that the problem of 'space errors' has nothing in common with the problem of 'time errors' in this method. He assumed, and falsely, that since the Ross list had not been arranged according to his criterion, it could not be so arranged. He then derived lists by trial-and-error which he assumed, again falsely, to be different from the Ross lists. He had, indeed, empirically derived a Ross list and then balanced it according to the scheme shown in Table I. It is not surprising that his lists turned out to be 'as good, or bad, as those obtained by Ross.'

III. Having made the two mistakes demonstrated above, it is

not strange that Wherry should make a third in assuming that no rational solution for obtaining orders which satisfy both his 'space error' and 'time error' criteria has been suggested. He says:

> "It seems to the writer that the question of optimum lists is still an open one. So far he has been unable to find a rational scheme for bettering all such lists, but surely, if it can be achieved empirically for seven variables (the writer has also bettered the Ross list for five variables in the same manner), it can be done rationally for any number. The search for such a rational method should not be stopped due to a belief that Ross has achieved its solution."

This brings us to a consideration of just what Wherry means by and demands of a 'rational method.' It seems to the writer that it is a sufficiently rational method to obtain the order sought from the matrix proposed by Ross and then to arrange the pairs within the list to satisfy the Wherry criteria. But Wherry implies that he expects something more of a rational method; one gathers that until a matrix has been devised such that one can read from it a list of orders which satisfy both the 'space' and 'time' criteria, Wherry will continue to consider the question an open one. Such a matrix is shown in Table III.

TABLE III

REARRANGEMENT OF ROSS'S MATRIX TO SATISFY WHERRY'S CRITERION

I	II	III	IV	V	VI	VII	etc.
1–2	2–3	1–3	3–4	1–4	4–5	1–5	etc.
3–n	n–4	4–2	2–5	5–3	3–6	6–4	etc.
4–(n – 1)	(n – 1)–5	5–n	n–6	6–2	2–7	7–3	etc.
5–(n – 2)	(n – 2)–6	6–(n – 1)	(n – 1)–7	7–n	n–8	8–2	etc.
6–(n – 3)	(n – 3)–7	7–(n – 2)	(n – 2)–8	8–(n – 1)	(n – 1)–9	9–n	etc.
7–(n – 4)	(n – 4)–8	8–(n – 3)	(n – 3)–9	9–(n – 2)	(n – 2)–10	10–(n – 1)	etc.
8–(n – 5)	(n – 5)–9	9–(n – 4)	(n – 4)–10	10–(n – 3)	(n – 3)–11	11–(n – 2)	etc.
9–(n – 6)	(n – 6)–10	10–(n – 5)	(n – 5)–11	11–(n – 4)	(n – 4)–12	12–(n – 3)	etc.
10–(n – 7)	(n – 7)–11	11–(n – 6)	(n – 6)–12	12–(n – 5)	(n – 5)–13	13–(n – 4)	etc.
11–(n – 8)	(n – 8)–12	12–(n – 7)	(n – 7)–13	13–(n – 6)	(n – 6)–14	14–(n – 5)	etc.
etc.	etc.	etc.	etc.	etc.	etc.	etc.	etc.

This matrix is identical with that given by Ross (1, 379) except that the orders of all pairs excepting the first have been reversed in all of the odd-numbered columns. The rules for its use are the same as for the original matrix and are as follows:

Use $\frac{n+1}{2}$ rows and $n - 1$ columns. In the $\frac{n+1}{2}$ row the fixed number in the first column will be $\frac{n+3}{2}$ and the number on the

variable scale will also be $\frac{n+3}{2}$, so that at this point the number on the fixed scale will be identical with that on the variable scale. Similar identities will appear in all odd-numbered columns. The following rule governs the use of the $\frac{n+1}{2}$ row: In the odd-numbered columns (where the repetition occurs) replace the number in the variable column by one. In the even-numbered columns disregard the entry in this row altogether. These rules are the same as those stated in 1934.

A consideration of the properties of this matrix will indicate the characteristics and limitations of the lists which may be derived from it. In the first place, it may be well to consider whether the spacing is maximal; that is, whether the lists are optimal with regard to 'time errors.' It is obvious that if we have n items, we will have $n(n-2)$ spaces which separate pairs involving the same item. Since in this discussion n is always odd, it is obvious that if the items are arranged in 1–2, 3–4, 5–6, etc., order, the greatest possible spacing will be insured for the 1 and 2. The number of such complete pairs will be $\frac{n-1}{2}$ and there will be one item left over. With this extra item we must combine either a 1 or a 2. If we put in the 1, then the spacing between 1's is $\frac{n-3}{2}$, and the 2 appearing in the next pair will be separated from the 2 in the first pair by $\frac{n-1}{2}$ pairs. A moment's consideration will show that maximal spacing between pairs involving the same item will be maintained if the pairs be arranged in 1–2, 3–4, 5–6, etc., order until the single term is reached, combining the single term with one and continuing the series $n-1$, 2–3, 4–5, etc., until the final term $(n-1)$–n is reached, and then continuing the series in 1–2, 3–4, 5–6 order again. Of course, this order makes no attempt at presenting all the possible pairs, but it obviously has the characteristic of maximal spacing.

A study of this 'maximal' list will show that in its arrangement the spacing $\frac{n-1}{2}$ is maintained half the time and the spacing $\frac{n-3}{2}$ is maintained for the remainder of the time. Since this represents maximal spacing, any other order which maintains the same total of spacings between pairs obviously represents the optimal situation. It may be seen from an inspection of the matrix that the orders

obtained from it maintain the spacing $\frac{n-1}{2}$ for half the pairs and the spacing $\frac{n-3}{2}$ for the other half. It therefore must represent the optimal spacing obtainable for the list.

Having demonstrated the 'optimal' nature of the lists with respect to the 'time' criterion, it is now necessary to investigate their conformity to the 'space' criterion. An inspection of the matrix will show that it is impossible to reverse the order of any pair without causing four 'space errors' if the pair reversed is in the body of the table, and two 'space errors' if the pair is either in the first or last columns or a pair involving 1 as one of its items. Any attempt at compensating for these errors leads to changes in adjacent columns, these changes cause more errors, and a continued attempt at compensation results either in reversing the order of all the pairs in the table or of coming back to the original form. Since this characteristic of the matrix is determined by the *a priori* balance of the table, it follows that as it stands, the matrix gives the minimum number of 'space errors' and is therefore 'optimal' in that regard.

The next question concerns the number of 'space errors' which we may expect from the use of the matrix. A consideration of the nature of the matrix and the rule for its use will show that since at the point of identity in the first column the number in the variable column which was replaced by 1 was $\frac{n+3}{2}$, it follows that in the first column the variable number which precedes the 1 is $\frac{n+5}{2}$, the one before that $\frac{n+7}{2}$, and so on to n. A further inspection of the matrix will show that all variable entries are balanced *a priori*, with the exception of the first entries in the 3rd, 5th, 7th and remaining odd-numbered columns. Also it will be seen that the last entry on the fixed side of the third column is $\frac{n+5}{2}$, the corresponding number in the fifth column $\frac{n+7}{2}$, etc. But it is likewise obvious that to the left of any one of these terminal items, the item occurs only in the variable part of the matrix, and that the fixed and variable portions are balanced in the opposite directions. It therefore follows that in passing from one arrangement to the other, there is an unavoidable 'space error,' any attempt to correct which

by changing the order within any given pair increases the number of errors in the matrix as has been shown above.

It would appear, then, that the errors are inescapable and that they involve all numbers from $\frac{n+5}{2}$ to n, once each, in their combination with 1. A consideration of the first row of the matrix will show that any number m which is greater than three will find itself in an anamolous position in the $(2m-3)$ column. Now we know from above that all numbers greater than $\frac{n+5}{2}$ will be involved in 'space errors' in the last row of the table. We have only to investigate the numbers less than $\frac{n+5}{2}$ for any possible exception.

Since we use $(n-1)$ columns, and the first 'space error' for any given item occurs in the $(2m-3)$ column, all items less than $\frac{(n+5)}{2}$ for which $(2m-3)$ is less than $(n-1)$ will escape from 'space errors.' It can be shown that the first error for $\frac{n+3}{2}$ is in the n column, which lies outside the matrix, hence its error does not occur. For all items numbered less than this, the column of first error occurs within the matrix (excepting 1 and 2).

The items 1 and 2 appear uniquely in the matrix. It is obvious that in the arrangement of any series of numbers into lists such as those here considered it is possible to balance items 1 and 2 perfectly, inasmuch as they are dependent on each other only in one common pair and are restrained in their positions by no other conditions.

It follows from the above considerations that in any list of n items there will be $(n-3)$ 'space errors,' and that the three exceptions will be for items 1, 2, and $\frac{n+3}{2}$.

One other point is of note. From the above discussion it would appear that the item in error always appears paired with one. This is true if the lists are read in the order in which they appear in the matrix. If they are read in reverse order, the 'error' apparently moves. For example, in the list for seven items in Table I where the pairs 1–3, 1–4, 1–6, 1–7 were considered as those having 'errors,' if the list is read in reverse order we mark the pairs 2–3, 3–4, 5–6 and 6–7 as apparently involved in the error. It is necessary to keep in mind that in a strict sense no pair or item is involved in an *error* but that the 'error' is in their arrangement. Which pair is out of order is a purely relative matter.

From the above considerations we may conclude that a 'rational method' for finding optimal orders has been found; that this method is merely a restatement of Ross's method; that the matrix proposed gives 'optimum' conditions for 'time' and 'space' errors; that the optimal 'time' condition involves a separation of $\frac{n-1}{2}$ half the time and of $\frac{n-3}{2}$ the other half; and that the optimal 'space' condition has $(n-3)$ 'errors' involving all items except 1, 2, and $\frac{n+3}{2}$.

IV. Wherry's fourth point is that "one serious error of application was made by Ross due to his having overlooked a possibility of greatly multiplying the results of his discovery." Wherry then proceeds to demonstrate this 'multiplication of results' by writing the list for seven items in 56 different ways.

To obtain these 56 variant forms, Wherry first substitutes certain permutations of the first seven digits for the order 1, 2, 3, 4, 5, 6, and 7. The permutations chosen rest on the following principle:

> "It is obvious that the item labelled '1' could be any of the seven items involved, and so on for any of the remaining items. All that we must do is to maintain the same interrelationships among our items. Thus, if all of the items be shifted clockwise one position, we may make the substitution indicated in our series without disturbing any of the relationships involved."

Unfortunately, Wherry has at this point become so intent on 'space errors' that he has forgotten about 'time errors,' for when substitutions are made, the 'space' ordering remains the same, but the 'time' ordering changes. Thus in Wherry's empirical list, pairs involving the first item are separated by two pairs throughout the original empirical list, but in the first variant the first four items involving '1' are separated by two pairs and the last three by three pairs.

But if 'time' order is to be sacrificed, one wonders why Wherry, after criticizing Ross for not multiplying his results, stops at a meagre 14 fundamentally variant lists. There are obviously 5040 different ways of substituting for the original seven items (the permutations of seven things taken seven at a time being factorial seven), and when we reverse the order and the pairs, as Wherry suggests, we emerge from the somewhat laborious process with a total of 20,160 different lists in place of Wherry's 56. These 20,160

lists satisfy all of the criteria that Wherry's lists do with regard to 'space errors' and 'time errors.'

V. Wherry's last criticism is of Ross's suggestion that "if the orders are repeated in reverse arrangement directly after the experiment has been completed in the given order, the effects of fatigue may be balanced out." Wherry points out, and quite correctly, that if this method is used, optimal spacing between pairs necessarily breaks down at the mid-point of the double series.

Wherry's suggested solution of this unfortunate situation is cumbersome and equivocal. He first finds 14 lists which if presented in immediate succession alleviate the immediate proximity of the terminal items in the Ross list with the identical items presented as the first pair in the reversed list, and in which, over the 14 series, each pair appears on the average at the mid-list position. It is not clear in his statement whether he expects the subject to proceed continuously through the entire 14 lists, or whether he is to be allowed to rest between lists. The inference is that the subject rests, for Wherry suggests as an alternative that

> "if the number of subjects used is always taken as some multiple of $2n$, it will be possible to use each of the selected lists an equal number of times, even though each subject is asked to react to a single list, thus securing a balancing out of the fatigue factor."

But whether a rest is given between lists is crucial. If no rest is permitted, then it is essential that the second half of the procedure be the mirror image of the first half; this is necessary in all ABBA orders for balancing out fatigue or practice effects and a condition which Wherry's method violates. It then becomes a matter of opinion, and experimental verification, as to whether the fatigue balance or the 'time' balance is the more important.

If a rest is to be allowed, then the 'time' balance is broken by the Wherry method, and it seems most expedient to repeat the list as Ross has suggested in reverse order after the rest period.

If a group of subjects is to be used, each performing on a single list, it is obviously sufficient if half the subjects go through one list in forward order and the other half go through the same in reverse order. The criteria of balance for fatigue and practice are then satisfied.

It would appear that Wherry's criticism logically hinges on Ross's use of the term 'directly after.' The writer agrees that no harm

would be done by allowing a rest period at this point, and that such a period would serve to maintain the optimum nature of the Ross lists.

Reply to Wherry

Wherry has criticised the work of Ross with regard to the 'optimum' nature of the orders for the presentation of pairs in the method of paired comparisons as suggested by the latter in 1934. Below are given the points in Wherry's summary with their answers:

> "1. It is shown that Ross's optimum lists are not optimum in all senses thus reopening the search for a rational solution of this problem."

It has been shown in I that the criteria for an optimum list suggested by Wherry were not those applied by Ross in 1934, but that if they be applied to the Ross lists, the lists derived from the matrix may be made to conform to Wherry's criteria.

In III the original Ross matrix has been rewritten by the simple expedient of reversing all pairs except the first in the odd-numbered columns, so that lists which conform to Wherry's criteria can be read directly from the matrix. This is, presumably, the rational method sought by Wherry.

It has been demonstrated in III that such lists are optimal with regard to both 'spacing' and 'timing.'

> "2. An empirically derived list for seven items is presented which is superior to the list given by Ross."

It has been shown in II that Wherry's empirical list is the corresponding Ross list in which the numbers 5, 4, 6, 3, 1, 7, 2 have been substituted for the Ross numbers 1, 2, 3, 4, 5, 6, 7 respectively, the Ross list then turned upside down, and the 'time' criterion of Wherry applied to individual pairs so derived.

> "3. A method is given whereby any list arrived at either rationally or empirically may be rewritten in $8n$ different ways."

It has been shown in IV that the Ross lists may be written in four times factorial n different ways. This means that for the list of seven items there are 20,160 different ways of writing the list without changing any of their optimal qualities. Wherry's selection of 56 of these seems to be a methodological accident.

> "4. It is shown that $2n$ of these lists may be combined in such a fashion that fatigue effects, introduced by the use of a single list, are cancelled out."

It has been shown in V that the lists proposed by Wherry do not satisfy the criteria for fatigue balance if they are given in a continuous series, and that the method is unnecessary if rest periods are introduced.

> "5. It is suggested as an alternative to the above that the number of subjects selected be some multiple of $2n$, and that each list be used in its proper order an equal number of times, each subject doing just one list."

It has been shown in V that if lists are to be given to a group of subjects it is sufficient to let half the subjects go through one list, and the other half go through the same list in the reverse order.

Summary

It has been shown that Wherry's empirical lists are variants of Ross's rational lists, hence there is no fundamental conflict between Wherry's proposals and Ross's earlier lists. One additional criterion, suggested by Wherry and not used by Ross, is easily fitted into the Ross matrix. This provides for an alternation of position from left to right as the same item recurs in subsequent pairs. Formal proof is given that the modification provides an optimal arrangement. Wherry's only remaining criticism is of Ross's suggestion that fatigue effects may be reduced by repeating the lists in reverse order. Wherry suggests a laborious alternative, requiring the use of $2n$ lists and some multiple of $2n$ subjects. It is shown that all that is necessary to satisfy Wherry's demands is to use one list, half of the subjects going through in one order, half in the reverse order. The number of possible lists which may be derived from Ross's matrix is far larger than suggested by Wherry. He suggests that there are $8n$ possible lists; actually four times factorial n may be constructed, each satisfying the criteria.

(Manuscript received April 5, 1939)

Bibliography

1. Ross, R. T., Optimal orders for the presentation of pairs in the method of paired comparisons, *J. educ. Psychol.*, 1934, **25**, 375–382.
2. Wherry, R. J., Orders for the presentation of pairs in the method of paired comparisons, *J. exper. Psychol.*, 1938, **23**, 651–660.

A REPLY TO ROSS ON ORDERS IN PAIRED COMPARISONS

BY ROBERT J. WHERRY

University of North Carolina

On the whole I quite agree with the present Ross paper. The modification of his matrix to yield really optimal lists rationally is just the solution for which I urged further search.

In this connection, since Ross seemed to be in doubt as to my use of the word rational, let me say that my understanding of this term was that it implied that the matrix itself would yield *final* results which would not be in need of any extended reshuffling of the sort referred to by Ross as 'balancing' in his original article. That such balancing may lead the 'juggler' to a 'fall' was evidenced by the fact that the Ross 1934 end-results were not so perfect (in the sense that fewer criteria were satisfied) as are his new lists taken directly from his *new* matrix.

That I had a perfect right to expect that *final* results of this degree of validity could be obtained *directly* from *some* matrix is shown by the fact that Ross has now rewritten his matrix to this end, thus incorporating all of the criteria into the schema instead of allowing the user to *stumble* on the correct arrangement.

As to my failure to recognize my empirical list as a Ross list I have no apologies to offer since it was *not* a 1934 Ross list. To be sure the 'time errors' are identical in both cases as I demonstrated, and it is therefore possible, by methods which Ross used after I suggested them, to substitute numbers in my list as Ross has done thus obtaining a list composed of his pairs. Such a list is not, however, an old Ross list either as it came from the 1934 matrix or as he had at that time reshuffled it, but rather agrees with the new criteria which I advanced, *i.e.*, trial and error shuffling was still needed to satisfy my added criterion.

While I might have further 'balanced' his list as he *later* did (see his Table I), I did not do so, since he had already admitted that his matrix results needed balancing and I had

shown that the balancing he had indulged in had not yielded results meeting all *possible* criteria, thus not meeting my conception of rational as defined above nor my concept of optimum. I must admit that had I done so I might have had the honor of modifying his matrix myself rather than simply leading him to do so.

Ross's attempt to claim that in my systematic rearrangement of lists I lost sight of time errors can most charitably be attributed to emotion at my attack on his perfect (?) system. What he does is to show that the 'time distribution' is no longer the same for item No. 1 as it was before. Anyone could have foretold this fact by noting that item No. 1 on first rotation of numbers has now taken the position formerly occupied by item No. 7 and so is now possessed of the time arrangement formerly belonging to that item. True this may happen to be a worse arrangement than before, but the overall distribution of time arrangements is of course left unchanged, some items being worse while others have become better. The total arrangement is always of seven types as shown in tables I and III of my article.

Why Ross should in one sentence claim that since "having made the two mistakes demonstrated above, it was not strange that Wherry should make a third in assuming that no rational solution for obtaining orders which satisfy both his 'space error' and 'time error' criteria has (*sic*) been advanced," and then try to maintain this fiction of my error by showing that the *new* matrix which he proposes is 'identical with that proposed by Ross except' that it is different, I cannot see. In my former article I praised his rational solution for the single criterion of 'time errors.' I am now willing to admit that his *new revised* matrix satisfies all criteria suggested to date. Personally I think he should thank me for leading him to improve his system, rather than attacking me for saying his system was not perfect when it was not.

As to the question of fatigue elimination I must partly agree and partly disagree. Since he has confessed the sins (violation, by his immediate reversal of his list, of all of the criteria) I cannot disagree with that part of his paper. On practically every other point I find myself still disagreeing.

The reason that I did not derive over 5000 lists in place of 14, each to be written in four manners, should have been apparent to Ross. I was not deriving these for amusement, but for the purpose of overcoming the error he had introduced when balancing for fatigue effects. It was my purpose then to do two mutually exclusive things: (1) to find lists which could be joined together without introducing errors of any kind, and (2) to have each pair of items occupy as large a *variety* of positions in the list as possible while averaging exactly mid-position. The method of unitary rotation which I used, rather than random substitution, as Ross suggests, was best adapted for securing this second condition. That I achieved both of these objectives—and therefore stopped the development of new lists—was demonstrated in tables VIII and IX of my article.

If it be 'equivocal' to suggest alternative procedures then I suppose I was guilty of this. What I tried to indicate was that my first preference was for the single subject immediately—without rest—to complete *all* of the lists at one sitting. I have had no reason for changing my opinion on this matter.

Just why a mirror image should be necessary when time intervals are allowed is not clear. True enough as Ross says any ABBA arrangement demands this, and that any such system (sufficient time having been allowed to make all of our other rules inoperative?) will successfully meet the criteria for balance for fatigue and practice, which is no more than to say that each pair will have appeared on the average in the mid-position. Why he should charge that my system of balancing has failed to achieve this after he had correctly stated above that this condition was met, I cannot understand. Perhaps it is a sample of worship of a sacred cow (mirror image satisfaction of the criterion) without a realization of just why the cow is sacred?

That my method was more 'cumbersome' I did not nor do not deny. It seemed, however, to offer an improvement over the mirror image method, in that while achieving the 'each item on the average at the mid-point' criterion, it widely

varied the position of the pair from list to list. One surely has the right to doubt whether a pair judged by half of the persons in position 2 and by the other half in position $n - 1$ has really been equated to another pair which for all judges has been exactly in the $n/2$ position. While my 14 lists met the criterion of mid-point average position they in addition allowed 14 of the pairs to vary in actual position from 1st to 21st, while the other 7 pairs varied from 2nd place to 20th. In addition to this scatter of range, 7 of the 21 pairs showed no duplication of position, 7 more pairs show but 1 duplication of appearance—a double appearance in the 11th-midpoint-position, and the remaining 7 pairs show two duplications of position at the 4th and 18th positions (equidistant from the ends). While this is not perfect it is at least an improvement over the mirror image concept.

Another virtue of this method over the mirror image technique is that this variation through the 14 lists permits what Ross mistakenly assumed to be an error, *i.e.*, the fact that different ones of the seven items to be paired assume each of the seven necessary time error patterns two times. Thus for item No. 1 we have separations of 2–2–2–2–2 in lists 1 and 23R, 2–2–2–2–3 in lists 6R and 25, 2–2–2–3–3 in lists 2R and 22, 2–3–2–3–2 in lists 4 and 27R, 2–2–3–3–3 in lists 3 and 28, 2–3–3–3–3 in lists 7 and 24R, and 3–3–3–3–3 in lists 5R and 26. The same thing could be demonstrated for each of the other six items. Use of the 14 lists thus balances out any error due to this unavoidable difference in time pattern due to the use of a single list and its mirror image.

On the above grounds I still feel inclined to suggest the use of the 'cumbersome' extension of the Ross technique. I cannot then accept his statement that "it becomes a matter of opinion and experimental verification, as to whether the fatigue balance or the time balance is the more important." I feel that I have shown above that my extension is superior on *both* counts. It would be truer only to say that it remains a matter of opinion and experimental verification as to whether these improvements are worth the extra bother and labor.

In closing I must say once more that the new *revised* Ross matrix meets *all* of the criteria thus far proposed as well as they can be met in a single list. It, however, still involves *necessary* weaknesses, and I believe these weaknesses to be best overcome by the method of multiplication and combination of lists which I suggested. While these are absolutely necessary for a single subject working without rest, which would be the ideal condition if fatigue did not enter in, the advantages are still present in the alternative suggestion of using the proper combination of $2n$ lists, one for each of $2n$ equal groups of subjects.

(Manuscript received April 17, 1939)